Gamma Type Distribution: Maximum Likelihood Values of the T-Year Event and their Asymptotic Variance

by

Paul J. Pilon

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy

Department of Civil Engineering
School of Graduate Studies
University of Ottawa

ISBN 0-315-60060-8

UNIVERSITÉ D'OTTAWA
UNIVERSITY OF OTTAWA

ABSTRACT

Maximum likelihood and censored sample theory are applied for flood frequency analysis purposes to the Two Parameter Gamma, log Two Parameter Gamma, Pearson Type III, log Pearson Type III (LP3), and Generalized Gamma distributions. The logarithmic likelihood functions are given in terms of the fully specified floods, the historical information, and the parameters to be estimated. Solution of the appropriate transcendental equations yields maximum likelihood estimators of the parameters. T-year floods are expressed as a function of these parameters and the standard normal variate. The asymptotic standard error of estimate of the T-year flood is derived using the general equation for the variance of estimate of a function. The variances and covariances of the parameters are obtained through inversion of Fisher's information matrix. The method is illustrated by application of the LP3 distribution to two sites having historical information.

Monte Carlo studies were conducted for the LP3 distribution to analytically verify the accuracy of the derived asymptotic expressions for the 10-, 50-, 100-, and 500-year floods. Results indicated that the asymptotic expressions were accurate for both Type I and Type II censored samples, while the bias was less than 2.5%. Subsequently, the Type II censored data were subjected to a random, multiplicative error. Results indicated that historical information contributes greatly to the accuracy of the estimate of the 100-year flood even when the error of its measurement becomes excessive. It is demonstrated that historical information can significantly reduce the standard error of estimate of flood quantiles.

ACKNOWLEDGMENTS

The author wishes to acknowledge the supervisor of this research work, Professor K. Adamowski, for his personal encouragement throughout his academic career.

The author wishes to thank his wife Janice and his daughters Stephanie and Sarah for their personal sacrifices and understanding during the preparation of this thesis.

During the course of the research, the external research advisor, Mr. Robert Condie, passed away. His leadership, advice and friendship were greatly appreciated and are sorrily missed.

Environment Canada is acknowledged for financial assistance, provision of computing resources and for the typing of this document. Thanks are given to the Word Processing Unit for their excellent work.

TABLE OF CONTENTS

TABLE OF CONTENTS (Cont'd)

LIST OF TABLES

LIST OF FIGURES

NOTATIONS

a	scale parameter of the 2PG, L2PG, P3, LP3, and GG distributions
b	shape parameter of the 2PG, L2PG, P3, LP3, and GG distribution
cdf	cumulative density function
CFA	Consolidated Frequency Analysis Package
CS	coefficient of skew
CV	coefficient of variation
D	the determinant of the inverse dispersion matrix
E()	the expected value of ()
$F(t_c)$	the probability that $t < t_c$
F(x)	the probability function of x
$F(x_c)$	the probability function that $x<x_c$
F(y)	the probability function of y
$F(y_c)$	the probability that $y<y_c$ $= \frac{1}{\Gamma(b)} \int_0^{y_c} y^{b-1} e^{-y} dy$
$f(t_c)$	the probability density of t_c, an N(0,1) variate
f(x;a,b)	the probability density function of the distribution with parameters a and b
f(y;b)	the density function of y, a gamma variate parameter b $= \frac{y^{b-1}}{\Gamma(b)} \exp(-y)$
$f'(y_c)$	the first derivative of $f(y_c)$ $= f(y_c) \left[\frac{(b-1)}{y_c} -1 \right]$
ft^3/s	cubic feet per second $1\ ft^3/s = 0.0283\ m^3/s$

GEV	generalized extreme value
GG	generalized gamma
GT	algorithm by Cheng and Feast (1980) to generate gamma variates
G1	Gumbel I or extreme value type I
h	shape parameter of the GG distribution
k or n_c	the number of censored floods below the threshold x_c
L	the maximum likelihood function
LN	lognormal
LP3	log Pearson Type III
L2PG	log two parameter gamma
ln L	the logarithmically transformed likelihood function
ln()	the Naperian logarithm of ()
ML	maximum likelihood
m	- location parameter of the P3 and LP3 distributions - rank in descending order of magnitude
N(0,1)	the normal distribution with zero mean and unit variance
n	the number of fully defined floods
n_a	the number of fully defined floods above or equal to the threshold x_c
n_b	the number of fully defined floods below the threshold x_c
n_c or k	the number of censored floods below the threshold x_c
P3	Pearson Type III
pdf	Probability density function
Q_t	T-year flood
q	the ratio $n_c/(n_b+n_c)$
RMS	root mean square
r	the number of censored values greater than x_u

$r_{i,j}$	element i,j of a matrix
S	unbiased standard deviation
S_y	unbiased standard deviation of y
T	return period
t	- a random variable from an N(0,1) distribution - Student's t in equation 4.5
t_c	t evaluated at the censoring threshold
u_i	pseudo-random number from a uniform distribution over the range (0,1)
v	the number of u_i's in the generation of a pseudo-random standard normal deviate, t.
x_c	the censoring threshold
x_l	lower censoring threshold
x_u	upper censoring threshold
$\bar{x}$	arithmetic average of x series
YT	the total time span
y	the transformed variate
y_c	the transformed censoring threshold
$\bar{y}$	the arithmetic average of the transformed series
2PG	two parameter gamma
3LN	three parameter lognormal
$\Gamma()$	the gamma function of ()
γ_1	skewness
γ_2	kurtosis
$\gamma_{1,y}$	skewness of y
$\gamma_{2,y}$	kurtosis of y
$\psi()$	the psi or digamma function of ()

$\psi^1()$	the trigamma function of ()
χ^2	chi-squared
Σ	summation
ν	degrees of freedomm

CHAPTER 1

INTRODUCTION

1.1 Motivation

Various methodologies have evolved to fulfill the need to estimate frequencies of floods including both parametric (Condie et al. 1981; Hydrology Subcommittee 1982; Pilon et al. 1985a) and non-parametric (Adamowski 1985) approaches. It is common practice for engineers to obtain design floods by associating probability with values of streamflow discharge at a site of interest. The basic approach to obtaining the probability-discharge relationship is point, or single-station, frequency analysis. When deriving the estimated flood quantile magnitude, Q_T, it is usually considered desirable to have an estimate of its accuracy (Cunnane 1987).

The most common question which arises from this problem is how best can Q_T be defined, given the gauged record at the site. It is recognized that the gauged record contains a limited amount of information due to its inherent characteristics and the number of years contained in the sample. Hence, any additional information which can be used to supplement the gauged record and which leads to increases in the accuracy of the estimation of Q_T would be advantageous.

Leese (1973a,b) introduced to hydrologists the ability to estimate the parameters of a distribution describing annual maximum flows from a river for which censored data existed. Censored data represents information supplemental to the gauged record and is considered non-standard data in classical flood frequency analysis. She demonstrated the marginal value of the approach by its subsequent reduction in the standard errors of the estimates of T-year floods for various values of T. The inclusion of non-standard data resulted in an increase in the accuracy of the estimate of the flood quantile, Q_T. She described two commonly occurring non-standard data: missing peaks in continuous chart records and historically marked flood peaks.

Leese's contribution to hydrology was partly due to the work of several statisticians regarding the rigorous treatment of censored samples (Hald 1949; Cohen 1950; Harper and Moore 1968). Hydrologists, like statisticians, want to make the best use of what little data are available. However, it was not until the early 1980's that a small flourish of activity could be perceived in the hydrologic literature. Recently, more studies are being published, and the approach is rapidly finding inroads especially in paleohydrologic and hydroclimatic change studies (Liebscher 1987; Baker 1987a).

Recent work (Cohn 1986; Stedinger and Cohn 1986a,b; Hosking and Wallis 1986a,b; Pilon et al. 1987) has demonstrated that the treatment of the censored sample using maximum likelihood theory can greatly increase the precision of parameter estimation methodologies leading to an ultimate increase in the accuracy of the estimation of the flood quantiles. This approach using one of the more common frequency distributions - the log Pearson Type III (LP3) - has not yet been achieved, due perhaps to its mathematical complexity and the need for extensive computer capabilities. However, its realization would be of great theoretical and practical value.

The motivation of this study is to develop the approach for the LP3 distribution and derive its asymptotic variance. This will lead to demonstrations of the worth of non-standard data in the reduction of variance.

It should be recognized that the LP3 distribution can easily be shown to fall in the gamma family of distributions. Several members of this family are being used in frequency analysis. They are the two parameter gamma (2PG), the log two parameters gamma (L2PG), the Pearson Type III (P3), and the generalized gamma (GG) distributions. Extension of the approaches to these distributions is as well important and will be achieved.

1.2 Literature Review

1.2.1 Flood Frequency Analysis

Flood frequency analysis relates the magnitude of discharge with its probability of occurrence. This relationship is usually accomplished by use of a cumulative density function. The presence of historic highs in the form of historic information, low outliers, or zeros will complicate the analysis. Eight combinations are possible:

(a) The standard case; alone or in conjunction with

(b) Historic highs

(c) Historic highs and low outliers

(d) Historic highs, low outliers, and zeros

(e) Historic highs and zeros

(f) Low outliers

(g) Low outliers and zeros

(h) Zeros

All cases except (d) and (e) have been found in hydrometric records from Canadian rivers.

The flood frequency analysis program (Condie et al. 1981) commonly used in the National Flood Damage Reduction Program of Canada does not handle the aforementioned non-standard cases. This program provides design flood estimates and their standard error for Gumbel I (G1), lognormal (LN), three parameter lognormal (3LN), and LP3 distributions. Parameters of the distributions are estimated using maximum likelihood theory, while moment estimates are as well given for the LP3 distribution. However, this program handles only the standard case which is only one of the eight combinations presented above.

In recognition that many hydrometric sites could be considered non-standard samples, a second package is available (Pilon et al. 1985a). It includes the generalized extreme value (GEV), 3LN, LP3, and Wakeby distributions. The methodology for the treatment of zero flows, low outliers, and historic floods is given by Pilon et al. (1985a) and Pilon et al. (1985b). Historic information is included in the estimation of the parameters through the use of historically-weighted moments for the GEV and LP3 distributions. Maximum likelihood and censored sample theory are used to derive the parameters of the 3LN distribution (Condie and Lee 1982). Regression of observed floods on empirical exceedance probabilities is used to estimate the Wakeby parameters in the historical case.

The new package differs from the old in that the standard errors of the flood quantiles are not computed. At the time, the theoretical derivations did not exist for the computation of the standard error in the historical case for any of the distributions contained in the new package. This is unfortunate, as the standard error is commonly used to assess the accuracy of a particular quantile estimate. The accuracy of at-site information is also an important input in the evaluation of hydrometric networks using generalized least squares (Moss et al. 1985; Thomas and Cheng 1985).

The most commonly used method for the inclusion of historic information in the estimation of parameters is historically-weighted moments. The U.S. Water Resources Council recommended its used in Bulletin 17 in 1976. Thomas (1985), reviewing the application of historically-weighted moments, ascribed the procedure to Mr. Fred Bertle, formerly of the U.S. Bureau of Reclamation. Thomas considered the inclusion of this procedure to be one of the most significant modifications to Bulletin 15. However, he noted that several studies have demonstrated that censoring theory and maximum likelihood estimates are more efficient but computationally more intensive. Cohn (1986), Stedinger and Cohn

(1986a), Condie and Lee (1982), and Pilon et al. (1987) have shown clearly that historically-weighted moments are less efficient than maximum likelihood based procedures. In addition, the maximum likelihood process lends itself to a theoretical derivation of the standard error of the quantile. Thus, this approach will be adopted in the present study.

The method of maximum likelihood is attributed to Sir Ronald Fisher. He first introduced the approach in 1912 (Fisher 1912), but presented a more complete exposé in 1922 (Fisher 1922). He showed how parameters of a distribution could be obtained as well as how to determine the asymptotic estimates of the parameters' variances and covariances. This led to the asymptotic variance of estimate of the quantiles, the square root of which is referred to as the standard error of the quantile which is sometimes referred to as accuracy. He demonstrated (1922) that the maximum likelihood procedure gave more efficient results than the method of moments for the P3 distribution.

In summary, censoring theory coupled with the maximum likelihood approach constitutes an important and powerful methodology which will be further developed in this study for the analysis of the frequency of floods.

1.2.2 Censoring Theory

In order to increase the accuracy of the estimation of quantiles, it is necessary to include information beyond that which is contained in the standard or classical streamflow record. Historic information is observed since its magnitude exceeds some threshold of perception. In statistical terminology a record of floods whereby this information can be extracted is termed a censored sample.

Therefore, censoring depicts a property of the sample. That is, if n values of x are observed between some lower threshold, x_l, and some upper threshold, x_u, and it is known that a certain number of values, say r, are not observed but are known to be larger than x_u, then the sample is said to be "censored on the right" at x_u. Conversely, "censoring on the left" occurs when it is known that a certain number of observations (and not their exact magnitudes) are less than x_l. In addition, "double censoring" exists when a known number of values are below x_l and another group are above x_u. There would remain n fully defined values and a number of values known to be less than x_l and a number greater than x_u.

In this scenario, a number of values are known to have occurred below or above some fixed value. That is, the censoring took place at a certain fixed point, the magnitude of which does not vary. This is called Type I censoring (Kendall and Stuart 1979, p. 551). This is analogous to the out-of-bank flood situation in flood hydrology and could be considered a sample censored on the right.

A second type of censoring is defined when a fixed proportion of the observations occur below x_l or above x_u. This is termed Type II censoring (Ibid, p. 551) and is easily demonstrated in industrial quality control. The variate under study is usually time, such as used in mean-time to failure experiments. It may be understated that in such studies one does not usually want to have all products fail in order to assess the statistic, as one may have to wait an indefinite long period of time. Thus, an underlying distribution is assumed and after a certain portion of items have failed, censoring theory can be used to yield the mean-time to failure of the product. This is termed Type II censoring, and it usually is "censored on the right".

In summary, it may be stated that in Type I censoring, the number of occurrences less than x_l or above x_u is random. The Type II censored sample is different in that the number of censored items is usually fixed a priori. Theoretical expectations and asymptotic statistics are derived in this study using the Type I model. It is usually assumed that the asymptotic results of both are analogous, but this is not necessarily so as this has not been analytically proven (Ibid, p. 552). Monte Carlo type experiments will be performed with the purpose of showing that Type I derived statistics are applicable in the Type II case.

1.2.3 Sources of Historic Information

The conventional sample available for flood frequency analysis consists of a series of maximum flows obtained from a continuous record of discharge at a hydrometric station. These maxima may be either instantaneous peaks or daily flows occurring in each calendar or water year, or in some specified season throughout the period of record. If the period of operation of the hydrometric station is n years, then the sample size is n, and n is typically quite small. Intuitively, any historic information which effectively enlarges n will improve the accuracy of the estimate of the design flood.

On many rivers some large floods may have occurred in the past, often many years prior to the installation of the hydrometric station. If their magnitudes and years of occurrence are known, then that historic information can be incorporated in the frequency analysis using this approach. Listed below are the definitions of symbols which will be further explained using the following hypothetical example.

If n_a = number of fully specified floods above the threshold

n_b = number of fully specified floods below the threshold

n_c = number of censored floods below the threshold

n = number of observed or fully specified floods

YT = total time span in years

then $n = n_a + n_b$ and $YT = n_a + n_b + n_c$

Let the example hydrometric record be available for the years 1940 through 1982. A flood of known magnitude occurred in 1920 and is known to be the largest since 1900. At first glance, the data available for analysis are the 1920 flood and the recent series from 1940 onward. There is, however, the additional information that in the 39 missing years - 1900 through 1919 and 1921 through 1939 - the annual flood was less than the 1920 value. These missing years are the censored data and the censoring threshold is the value of the 1920 flood. So $n_a = 1$, $n_b = 43$, $n_c = 39$, $n = 44$, and $YT = 83$. When performing an historical frequency analysis, only YT and the censoring threshold need be specified supplementary to conventional data requirements.

The term historic need not apply only to floods which occurred before the installation of a hydrometric station to collect a continuous record. Methods used are equally adaptable to the case of the occurrence in the gauged record of the largest flood or floods in the history of the area, provided that there is reliable local information that the flood or floods were the largest since some known date. Suppose that at the same hypothetical location, the 1960 flood is reliably known to be the second highest since 1900. Then $n_a = 2$ and $n_b = 42$; n_c, n, and YT remain at 39, 44, 83, and 1, respectively, but the censoring threshold then becomes the value of the 1960 flood.

The above described example is considered as an "out-of-bank" type of flood record and is frequently encountered in practice. For example, Chen et al. (1974) discuss the historical monuments placed in the Yangtze River over several centuries to document severe floods. Leese (1973a, 1973b) uses marker data in her analysis of the Avon River at Bath in England.

Usually, floods which rose above a certain level, generally bank full, were marked by dated stone markers and discharges have since been estimated, and the assumption is that in the years for which no markers exist, the flood was less than bank-full discharge. In an analysis of this type, the total time span, YT, is obtained from the data of the earliest marker, and the censoring threshold is bank-full discharge. The number of fully defined floods, n, is clearly the number of annual maximum floods in the gauged record plus the number of markers.

The above examples of sources of historic information represent the "highest within living memory" (Natural Environment Research Council 1975, p. 215) and the out-of-bank case. The first is a Type II censored sample, while the second example is of a Type I. The Type I example is statistically more tractable; however, the Type II scenario appears often and is an important case in flood hydrology.

Sources of historic information can be quite varied. Potter (1978) describes in detail the methodology a researcher can use in the United Kingdom to obtain reliable historical information. His sources are primarily recorded public information dating from the present to events prior to the Norman conquest. In the U.S., Thomson et al. (1964) document the occurrence of hundreds of floods in New England from the early seventeenth century to modern times. In comparison, the People's Republic of China has embarked on a national survey of historical flood markers in order to obtain information for the design of major dams (Zheng 1987; Shi et al. 1987).

Recently, paleohydrological methods have emerged as a major source of flood data which can potentially be analyzed as a censored sample (Costa 1978; Baker 1983; Lane 1986; Baker et al. 1986a,b; Hupp 1986; Baker 1987a,b; Liebscher 1987). Costa (1984) defined paleohydrology as "the study of the movements of water and sediment in channels before the time of continuous hydrological records or direct measurements". Baker (1987b) noted that this includes both historical floods predating stream gauging and modern floods which have not been directly observed by humans. Geomorphic evidence of past floods can be deduced from botanical studies of floodplain vegetation (Sigafoos 1964; Helley and LaMarche 1973; Hupp 1986) and stratigraphic geology.

The botanical approach to record extension is primarily through the analysis of damage to trees located in the floodplain. Trees continue to grow, even after extensive damage. This continued growth preserves the information and provides a chronometric tool for the dating of evidence. Sediments deposited in the floodplain can possibly be dated from tree ring analysis - dendrochronological techniques; bending and orientation of the tree; and sprout conditions (Sigafoos 1964; Helley and LaMarche 1973). Seedlings usually form on sediment deposits after 5 years and can be used to determine the period when the sediments were deposited (Helley and LaMarche 1973).

Botanical methods can provide information dating to the age of the trees in the floodplain. This is usually limited at the extreme to a few centuries. Stratigraphic surveys coupled with modern radiocarbon dating procedures can extend the period of investigation to several millennia (Baker et al. 1979; Kochel and Baker 1982; Xu and Ye 1987).

The study of slack-water sediments, silt lines, and flood scars provide information on dates of flooding and their stage. Kochel and Baker (1982) and Baker et al. (1979), as well as several other sources, indicate non-alluvial channels provide protected locations for deposition and that their subsequent preservation yield the most useful data. It is preferred that these channels be of confined bedrock thus providing a stable control with a relatively steep stage-discharge curve. Estimates of discharge can be obtained using various indirect hydraulic methods (Kochel and Baker 1982; Shi et al. 1987; Baker 1987). Inaccuracy of indirect discharge measurements is due to errors from various sources (Kirby 1987). Kirby (1985) estimates a root mean square error of 25% for discharges measured by the slope-area method, one of the more commonly applied procedures. Errors in estimating discharge can increase greatly when the channel is alluvial and the stage-discharge relation's slope is mild (Baker et al. 1979; Kochel and Baker 1982; Baker 1987b).

Stratigraphic paleoflood samples usually provide a chronological progression of worst flood cases. That is, flood stages are recorded only by deposition of their sediment when the flood's crest exceeds the site's overburden elevation. Radiocarbon dating is performed on small bits of organic material deposited on the sediment's top layer (Baker 1983; Baker 1987a,b). In essence, the paleoflood record constitutes a censored sample with a varying threshold exceedance. This sample would be analogous to that obtained by flood markers placed only when the flood stage exceeded that of a marked past level. This is similar to the "highest in living memory" case, but the memory is updated for millennia. However, additional censored information is available in the paleoflood sample. That is, prescribed levels may not have been exceeded for certain time periods. Inclusion of such information in a model of censoring is recognized as being different and as being more complex than the "highest in living memory" case (Leese 1973a, p. 1541). "In order to determine which type of censoring is appropriate to describe a given situation, and what information an historical record provides, it is necessary to understand the processes that may lead to the creation of historical flood records" (Stedinger and Cohn 1986b, p. 274).

It is apparent that significant efforts are being made by governments to obtain historical information primarily for inclusion in flood frequency analysis studies. It is unfortunate that in Canada the significance of the inclusion of historical information is not fully recognized.

1.2.4 The Factors Affecting the Utility of Historic Information

Several factors have been identified which govern the utility of historic information in the analysis of the frequency of floods (Hosking and Wallis 1986a). These factors include the selection of the parent distribution, the choice of parameter estimation methodology, the choice of the flood quantile, the length of the gauged record and its variability, the measurement error of the gauged record, the measurement error of the historic information, and the stationarity of the annual flood series.

In this study, the gamma family of distributions is selected as the parent where the parameters are estimated by the method of maximum likelihood. Emphasis is placed upon the LP3 distribution as it represents the most widely used distribution in flood frequency analysis. The analysis of the censored sample is performed using maximum likelihood theory as it has been shown to be the preferred approach in increasing the accuracy of parameter and quantile estimations. The quantiles studied are the .9, .98, .99, and .998. These correspond with the 10-, 50-, 100-, and 500-year floods, respectively. These represent the quantiles used in many engineering applications.

The nonstationarity of a flood series is attributable to changes in land use and to meteorological conditions which are not characteristic of the contemporary basin. The assumption of stationarity is particularly tenuous when the historic information dates several millennia due primarily to the possibility of climate change. The influence of land-use changes on the flow regime may be assessed if the nature and timing of such changes are known. However, the effect of nonstationarity on the estimation of parameters and quantiles in frequency analysis is beyond the scope of this study. Hence, in this work it is assumed that the series for which historic information exists is stationary. It is also commonly assumed when performing frequency analysis that the systematic record is free of error. In this study, the assumption of error-free systematic record is as well made.

It has been suggested (Hosking and Wallis 1986b) that the magnitude of an historical flood may not be as accurate as that of gauged flood records. This would be particularly true when the worst flood in memory occurs prior to the start of the systematic record. Errors may be in part attributed to unrecorded geomorphological changes in the river regime and to extrapolations of the stage-discharge relationship. Knowledge of the stability of a river bed and an evaluation of the rating points and history of a stage-discharge relationship at a gauging site can assist in the assessment of possible error. Historical discharge can also be estimated by indirect methods such as the slope-area method or unsteady flow modelling. These approaches can be used to compliment rating curve data and can assist in rating curve extension.

Hosking and Wallis (1986b, p. 1610) studied the influence of measurement error pertaining to the historic flood. They found for the GEV distribution that when the measurement error exceeded ±25% for the historic flood, "then there is no overall advantage in including it in the analysis rather than using ... the gauged records alone".

A Monte Carlo study will be performed to assess the influence of measurement error with regards to bias and accuracy of the quantile estimates. A random multiplicative error will be associated only with the worst flood in memory.

In summary, several factors influence the utility of historic information in the analysis of the frequency of floods. A Monte Carlo study will be performed to assess the influence of measurement error of the historic flood for the LP3 distribution.

1.2.5 Flood Frequency Analysis Using Censoring Theory

If a sample of x's, size n, is drawn from the postulated distribution, but with the parameters as yet unknown, the likelihood function L can be expressed in terms of the sample and the unknown parameters. This likelihood function L is the probability that all the members of the sample were drawn from the distribution, and the principle of maximum likelihood states that the unknown parameters should be chosen to maximize L.

Suppose that the magnitude of a sample member is unknown, but the sample member is known to be less than a certain value x_c, the censoring threshold. Such a member is called a censored member. Then, the probability that the sample member was less than x_c, and came from the postulated distribution, is the probability function evaluated at x_c, $F(x_c)$. By extension the probability that r sample members were less than x_c is $[F(x_c)]^r$. For a censored sample from the postulated distribution with parameters yet to be determined and since the likelihood function is a probability, L can be expressed in terms of the fully specified sample members, the number of censored values below the censoring threshold, and the distribution parameters. Maximizing L by taking partial derivatives with respect to each parameter in turn and equating them to zero gives a set of simultaneous equations, the maximum likelihood estimators, which when solved give maximum likelihood estimates of the distribution parameters. Once the parameters have been estimated, the floods of the required exceedance probabilities or return periods can be computed. Examples of this will be shown in detail in Chapter 2.

Asymptotic variances can, in turn, be obtained by solving Fisher's information matrix. This requires the derivation of the double partial derivatives of L with respect to each parameter in turn and, subsequently, finding their negative expectations. These values, once obtained, can be substituted into the general expression for the variance of a function. With some manipulations, the asymptotic standard error of estimate of the T-year event can be obtained. This will be demonstrated for various distributions in Chapter 3. Monte Carlo experiments will be performed to verify the performance of the above theoretical developments.

The above represents a brief grammatical exposé of the approach, which was initially developed by statisticians (Hald 1949; Cohen 1950). As previously mentioned, Leese (1973a,b) introduced the concept to hydrologists. She adapted the approach with the G1 distribution having both Type I censored data and a systematic gauge record. She developed expression for the asymptotic variance of the T-year event and analytically demonstrated the reduction in the asymptotic error of the estimate by inclusion of historic flood marker information. Leese (1973b) demonstrated for the Avon at Bath that the inclusion of historic information decreased the asymptotic standard error of the 10-year and 100-year floods by 39.4% and 35.3%, respectively. The Natural Environment Research Council (1975, p. 213-215) adopted Leese's work as an example of censoring below a known threshold.

The work of Condie and Lee (1982) represented the next known published work in hydrology regarding censored sample application with maximum likelihood theory. They followed the work of Leese (1973a,b); however, they applied the approach using the 3LN distribution. They felt that the G1 distribution lacked "flexibility" and was "not an entirely satisfactory model of the frequency of floods in Canada". They derived the necessary formulae and provided an analytical methodology sufficient to estimate the parameters of the 3LN distribution from censored samples. Application of their approach was given for a Type II censored sample - Irwell River at Adelphi Weir, England. In addition, they evaluated the ability of their model with the technique of historically weighted moments using a Monte Carlo experiment. It was performed using a Type I censored sample scenario. Results indicated that the maximum likelihood procedure was "substantially less biased in the absolute sense than those obtained by the more conventional method of historically weighted moments" (Condie and Lee 1982, p. 60). In general, they concluded that the maximum likelihood method seemed "preferable".

The asymptotic standard error of estimate for the 3LN distribution was later derived independently by Condie (1986) and Cohn (1986). Cohn (1986) performed a limited Monte Carlo experiment in order to determine the behaviour of 9 different estimation procedures, including both maximum likelihood and historically weighted moments. Two different lognormal populations were generated. The first corresponded to a coefficient of variation (CV) of .25 and a coefficient of skew (CS) of .76, while the second had a CV of 1.0 and a CS of 4. Censoring thresholds were set to exceedance probabilities of .01, .02, .05, and .10. In total, two groups of 2500 samples formed the basis of the experiment. Each sample was comprised of 50 systematic years of record and an historical record length of 200 years. Cohn (1986, p. 92) concluded that "the likelihood procedures are the most efficient for fitting the range of LN3 [3LN] populations that are likely to occur in practice". Cohn and Stedinger (1987) and Stedinger and Cohn (1986b) presented results similar to that of Cohn (1986).

Prior to their work on the 3LN distribution, Cohn (1984) and Stedinger and Cohn (1986a) reported on a comparison of maximum likelihood based procedures with historically weighted moments for the two parameter lognormal (2LN) distribution. They performed a Monte Carlo experiment consisting of 100 cases composed of systematic record lengths of 10 to 100 years in combination with historical records of 20 to 200 years. Censoring thresholds were set in the ninetieth to ninety-ninth percentile range. Results were similar to their later work on the 3LN distribution. That is, the maximum likelihood based procedures were "much more efficient than the adjusted-moment [historically weighted moment] method" (Stedinger and Cohn 1986a, p. 790). In summary, they found that the maximum likelihood procedures could "extract the equivalent of an additional 10-30 years of gauge record from a 50-year period of historic observation" (Ibid, p. 785).

Condie and Pilon (1983) developed a maximum likelihood procedure with censored data for the lower-bounded case of the LP3. An example, the Floyd River at James, Iowa, was given to demonstrate the methodology. However, the asymptotic standard error of the estimate was not derived.

Hosking and Wallis (1986a,b) evaluated the worth of historic information for the G1 and GEV distributions for Type I censored samples based on Monte Carlo simulations. They also investigated the influence of measurement error of the historic flood on the estimate of various quantiles through the use of a random, multiplication error. They considered only maximum likelihood based procedures as "recent research has clearly demonstrated that this approach [historically weighted moments] makes inefficient use both of gauged records (Wallis and Wood 1985) and of historical data (Stedinger and Cohn 1986a) and does not merit serious consideration in a scientific approach to flood frequency analysis" (Hosking and Wallis 1986b, p. 1607). Results indicated that the three parameter flood frequency distribution gave superior estimates of quantiles as compared with the two parameter distribution. In addition, when error was associated in the estimate of the historic flood, the two parameter distribution proved inefficient as compared with the GEV distribution. Thus, it would appear "historical information is of great value provided either that historical discharges are accurately estimated or that the flood frequency distribution has at least three unknown parameters" (Hosking and Wallis 1986b, p. 1611). Hosking and Wallis (1986a), using the GEV distribution, found that "even the largest gauged records gain substantially from the inclusion of historical information". These results were of particular importance with regards to the use of paleological data.

In summary, censoring theory and the maximum likelihood approach will be used to obtain parameter estimates. Fisher's information matrix will be solved in order to obtain variance estimates of the parameters. These, in turn, will be substituted into appropriate forms of the general expression for the variance of the function. This will yield the expressions for the asymptotic variance of the estimated quantile. So far, this approach has been applied in hydrology and statistics to simple cases. It has yet to be developed and applied for the more complex 3 parameter distributions such as the LP3. Therefore, in this thesis, the above approach will be extended for the gamma family of distribution, which includes the LP3 distribution.

1.3 Objectives

Based on the literature review, it is obvious that censoring theory with maximum likelihood is the most accurate approach for the inclusion of historic information in frequency analysis. However, only simple cases so far have been considered. In this study, these procedures will be developed for the more complex and most commonly used distribution in flood frequency analysis.

The objectives of this study are:

1) to formulate and develop procedures:

   (i) for the estimation of parameters using maximum likelihood and censoring theories for the following distributions: (a) the two parameter gamma; (b) the log two parameter gamma; (c) the Pearson Type III (three parameter gamma); (d) the log Pearson Type III; and (e) the generalized gamma;

   (ii) to determine the asymptotic standard error of estimate of the parameters and quantiles of the above distributions.

2) to investigate using Monte Carlo simulation methods:

   a) the accuracy of the censored model's asymptotic results;

b) the applicability of Type II censoring model to estimate the asymptotic error of Type I information; and

c) the influence of measurement uncertainty of historic information on the asymptotic error.

3) to investigate the properties of Monte Carlo data simulation methods.

CHAPTER 2

THEORETICAL DEVELOPMENTS OF THE ESTIMATION OF PARAMETERS

2.1 The Two Parameter Gamma Distribution

The probability density function (pdf) of the two parameter gamma (2PG) distribution is given by

$$f(x;a,b) = \begin{cases} (\exp(-x/a)x^{b-1})/(a^b\Gamma(b)), & x\geq 0,\ a>0,\ b>0 \\ 0 & \text{elsewhere} \end{cases} \tag{2.1}$$

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

$$\Gamma(b) = \int_0^\infty x^{b-1}e^{-x}dx \tag{2.2}$$

and $\Gamma(b) = (b-1)!$ for positive integer values of b.

The pdf of the 2PG distribution can assume various shapes. Figure 2.1 shows the influence of b given a constant "a" parameter of one. The distribution is reverse J-shaped when $b \leq 1$ and forms a unimodal shape when $b > 1$ having a peak at $[(b - 1)a]$. Figure 2.2 shows the influence of "a" given a constant b. The value of "a" does not alter the form of the pdf, but affects only its scale. Thus, "a" and b are termed scale and shape parameters, respectively.

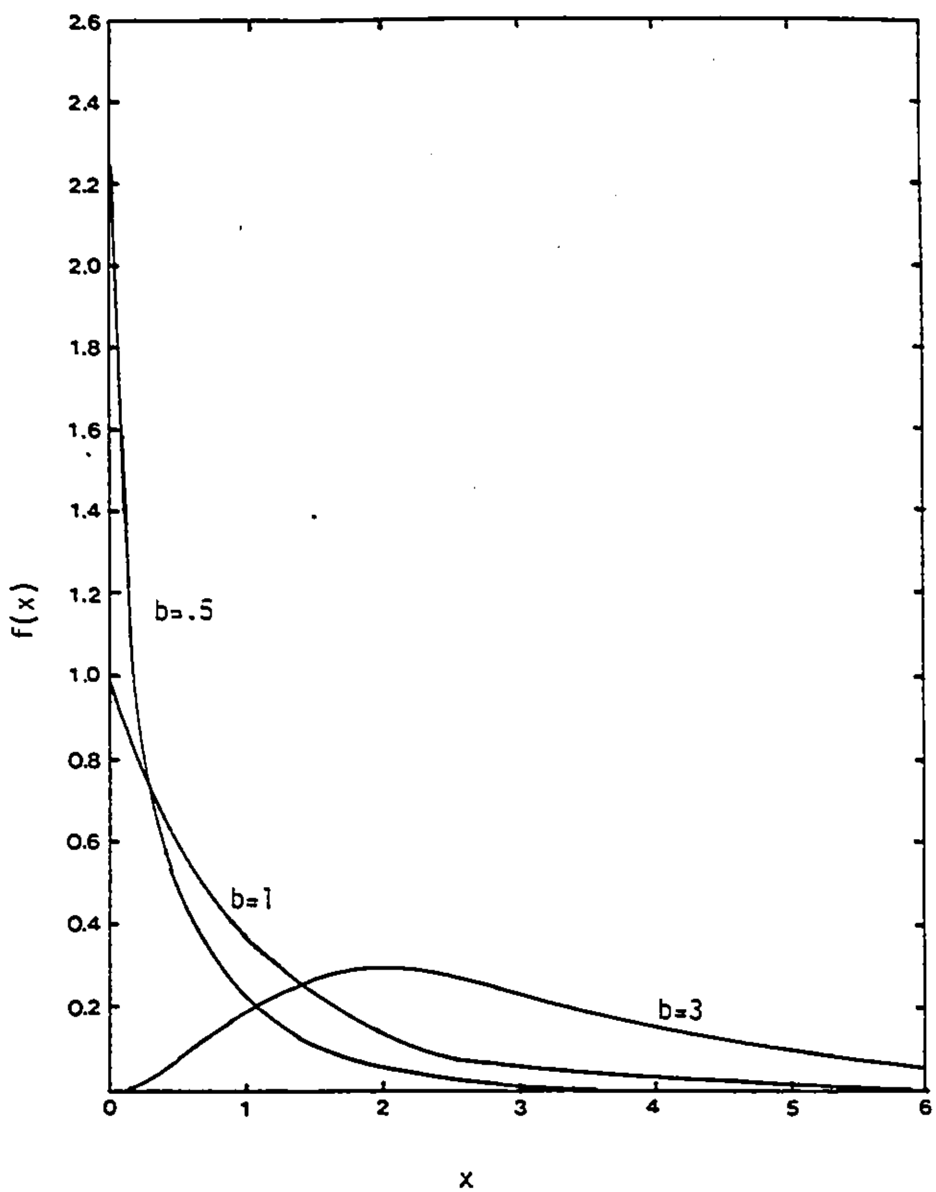


Figure 2.1: Shapes of the Probability Density Function of the 2PG Distribution Having a Scale Parameter, a, of One (Hahn and Shapiro 1967, p. 84).

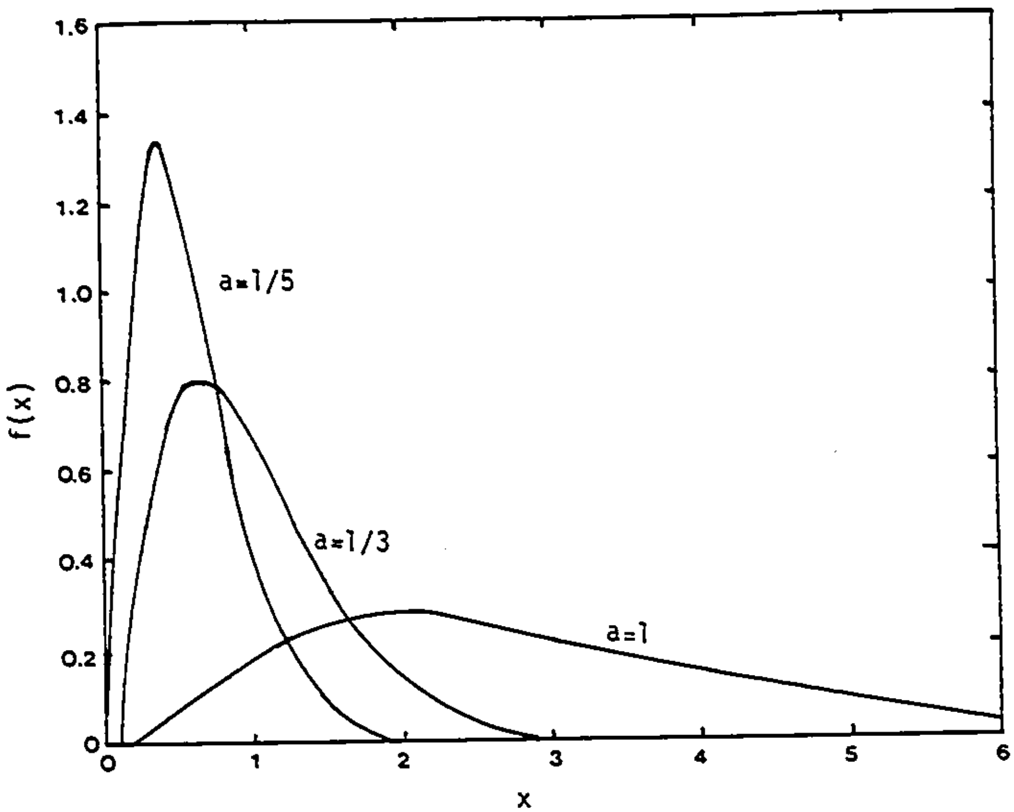


Figure 2.2: The Probability Density Function of the 2PG Distribution Showing the Influence of the Scale Parameter for a Constant Shape Parameter, b, of Three (Hahn and Shapiro 1967, p. 85).

Due to the wide variety of shapes of the 2PG distribution, it has seen numerous applications in engineering. These have varied from use in statistical queueing theory (Hahn and Shapiro 1967, p. 83), to the description of synthetic hydrographs (Croley 1980; Cruise and Contractor 1980), to drought analysis (Gueven 1983), to frequency analyses of annual precipitation and runoff (Markovic 1965) and extreme climatological data (Thom 1958).

Flood hydrology is most interested in the unimodal shape of the pdf associated with $b > 1$. It can be shown that the 2PG is a special case of the P3 distribution when its location or boundary parameter is set to zero. Thom (1958), Kendall (1966), Yevjevich (1972) and others have shown that estimates of the parameters can be obtained from the moment relationships:

$$\bar{x} = ab \tag{2.3}$$

$$S = a\sqrt{b} \tag{2.4}$$

where $\bar{x}$ is the sample's arithmetic mean and S is an unbiased estimate of the sample's standard deviation. The skewness, γ_1, and the kurtosis, γ_2, of the 2PG are related to b by

$$\gamma_1 = 2/\sqrt{b} \tag{2.5}$$

$$\gamma_2 = 3 + (6/b) \tag{2.6}$$

From these two equations, it can be seen that as b increases the skewness tends to zero and the kurtosis tends to three. The distribution approaches normality as b increases (Thom 1958, p. 118). However, the 2PG distribution is bounded at zero and unbounded above, while the normal distribution is unbounded below and above.

The coefficient of variation, CV, is defined as

$$CV = S/\bar{x} \tag{2.7}$$

Substituting equations (2.3) and (2.4) into equation (2.7) gives

$$CV = 1/b \quad \text{or} \quad \gamma_1 = 2CV \tag{2.8}$$

Algebraic manipulation of equations (2.5) and (2.6) give

$$\gamma_2 = 1.5\ \gamma_1{}^2 + 3 \tag{2.9}$$

Equation (2.9) and especially equation (2.8) can be used as a "first and rough test of goodness of fit for the two-parameter gamma function before a more efficient method of estimation (the maximum likelihood)" (Yevjevich 1972, p. 146) is applied. That is, when using moments to estimate the parameters of the 2PG, equations (2.8) and (2.9) should be tested. Deviations from equality can infer that the sample does not follow the 2PG distribution.

The logarithmically transformed likelihood function, lnL, of the 2PG distribution is given (Thom 1958; Kendall 1966) as:

$$\ln L = -nb\ln a - n\ln\Gamma(b) + (b-1)\Sigma\ln x - \frac{1}{a}\,\Sigma x \qquad (2.10)$$

where the summation is over the n sample values. The maximum likelihood (ML) equations are found by differentiating L and setting the result equal to zero. That is $\partial\ln L/\partial a = \partial\ln L/\partial b = 0$. Thus,

$$\frac{\partial\ln L}{\partial a} = \frac{-nb}{a} + \frac{1}{a^2}\,\Sigma x = 0 \qquad (2.11)$$

$$\frac{\partial\ln L}{\partial b} = -n\ln a - \frac{n\partial\ln\Gamma(b)}{\partial b} + \Sigma\ln x = 0 \qquad (2.12)$$

where $[\partial\ln\Gamma(b)/\partial b]$ is the digamma or psi function, $\psi(b)$. Simplification of equations (2.11) and (2.12) yields

$$\bar{x} = ab \qquad (2.3)$$

and

$$\psi(b) + \ln(a) = \frac{\Sigma\ln x}{n} \qquad (2.13)$$

Rearranging equation (2.3) and subsequent substitution into equation (2.13) yields an expression only in b:

$$\ln(b) - \psi(b) = \ln\bar{x} - \frac{\Sigma\ln x}{n} = A \qquad (2.14)$$

Equation (2.14) can be solved iteratively to find b. Once b has been determined, it is substituted into equation (2.3) to obtain "a". Thom (1958) developed an approximation which is more convenient when balancing (2.14) by hand. He gives b as the pertinent root of the quadratic expression

$$b = \frac{1 + \sqrt{1 + 4A/3}}{4A} \tag{2.15}$$

He notes that the value of b from (2.15) should be decreased by the value "Δb". The Δb is not readily expressed in mathematical form, thus Thom (1958, p. 119) and Markovic (1965, p. 9) give a table to estimate the values. The following expression can be used to closely replicate their tabular results.

$$\Delta b = .0015421 - .0123091(1/b)^{.57} + .026495(1/b)^{1.14} - .0065687(1/b)^{1.71} \tag{2.16}$$

where b ranges from .2 to 5.6. Δb should be set equal to zero for $b \leq .2$.

The above developments are applicable only to standard samples. Extensions to the non-standard sample follow.

2.1.1 Inclusion of Historic Information

For a sample size n of the variate x, and a further number k whose magnitudes are known to be less than the censoring threshold, x_c, the likelihood function, L, is given by:

$$L = [(n+k)!/k!][F(x_c)]^k \pi f(x;a,b) \tag{2.17}$$

where $F(x_c)$ is the integral of the pdf from zero to x_c. That is, $F(x_c)$ is the cumulative density function (cdf) of the 2PG distribution and represents the probability of a value being equal to or less than x_c. The cdf is:

$$F(x_c) = \frac{1}{a^b\Gamma(b)} \int_0^{x_c} e^{-x/a} x^{b-1} dx \tag{2.18}$$

The logarithmic likelihood function, lnL, is

$$\ln L = \ln[(n+k)!/k!] + k\ln F(x_c) + \Sigma \ln f(x;a,b) \tag{2.19}$$

A change of variate is used such that

$$y = x/a \tag{2.20}$$

and the density function of y is a gamma variate, parameter b, and is expressed as:

$$f(y;b) = \frac{y^{b-1}e^{-y}}{\Gamma(b)} \tag{2.21}$$

The cumulative density function becomes

$$F(y) = \frac{1}{\Gamma(b)} \int_0^{y} y^{b-1}e^{-y}dy \tag{2.22}$$

and

$$F(y_c) = \frac{1}{\Gamma(b)} \int_0^{y_c} y^{b-1}e^{-y}dy \tag{2.23}$$

where $F(y_c)$ is a form of the incomplete gamma function and represents the probability of $y<y_c$. Equation (2.19) can now be written as:

$$\ln L = \ln[(n+k)!/k!] + k\ln F(y_c) + \Sigma \ln f(y;b) \qquad (2.24)$$

From equation (2.20), dy/da = -y/a. The derivative of the censored term of (2.24) with respect to "a" is:

$$\frac{\partial \ln L}{\partial a} = \frac{k\partial}{\partial y}[\ln F(y_c)]\frac{\partial y}{\partial a} = \frac{kf(y_c)}{F(y_c)} \cdot \frac{\partial y}{\partial a} = \frac{-y_c kf(y_c)}{a\ \ F(y_c)} \qquad (2.25)$$

Combining equations (2.11) and (2.25) yields the first partial derivative of lnL with respect to "a":

$$\frac{\partial \ln L}{\partial a} = \frac{-nb}{a} + \frac{1\Sigma y}{a} - \frac{y_c kf(y_c)}{aF(y_c)} = 0 \qquad (2.26)$$

Combining equation (2.12) with the derivative of equation (2.24) with respect to b yields:

$$\frac{\partial \ln L}{\partial b} = -n\psi(b) + \Sigma \ln y + k[\partial \ln F(y_c)/\partial b] = 0 \qquad (2.27)$$

Parameters "a" and b are obtained by solving equations (2.26) and (2.27) simultaneously using numerical analysis methods. In equation (2.27), the terms $[\partial \ln F(y_c)/\partial b]$ can be evaluated by use of the work of Wilson and Hilferty (1931). They showed that $(\chi^2/\nu)^{1/3}$ is approximately normally distributed with a mean of $[1-2/(9\nu)]$ and a variance of $[2/(9\nu)]$. This gives (Condie 1977, p. 988)

$$X^2 = \nu\{1-2/(9\nu) + t[2/(9\nu)]^{1/2}\}^3 \tag{2.28}$$

where X^2 is the chi-squared distribution having ν degrees of freedom and t is the standard normal deviate at the required probability level. Substituting $2y = 2x/a = X^2$ and $2b = \nu$ and rearranging gives

$$y_c = [t_c/3b^{1/6}) - 1/(9b^{2/3}) + b^{1/3}]^3 \tag{2.29a}$$

or

$$t_c = [y_c^{1/3} + 1/(9b^{2/3}) - b^{1/3}]3b^{1/6} \tag{2.29b}$$

The Wilson-Hilferty transform has been shown to be accurate for skewnessess less than 3 and departs mildly at higher values (Kirby 1972). Thus, for practical applications, the correction is not applied.

Given that $F(y_c) = F(t_c)$, then by the chain-rule

$$\frac{\partial\,[\ln F(y_c)]}{\partial b} = \frac{\partial[\ln F(t_c)]}{\partial b} = \frac{\partial\,[\ln F(t_c)]}{\partial t}\frac{\partial t}{\partial b} = \frac{f(t_c)}{F(t_c)}\frac{\partial t}{\partial b} \tag{2.30}$$

The derivative of (2.29b) with respect to b gives

$$\frac{\partial t_c}{\partial b} = y_c^{1/3}/(2b^{5/6}) - 1/(6b^{3/2}) - 3/(2b^{1/2}) \tag{2.31}$$

Thus, $[\partial\ln F(y_c)/\partial b]$ can now be evaluated.

2.2 The Log Two Parameter Gamma Distribution

The pdf of the L2PG distribution is given by

$$f(x;a,b) = \begin{cases} (\exp[-(\ln x)/a]\ln x^{b-1})/(a^b x\Gamma(b)), x \geq 0, a>0, b>0 \\ 0 \quad \text{elsewhere} \end{cases} \quad (2.32)$$

The logarithmic likelihood function, lnL, corresponding to the pdf given by equation (2.32) is

$$\ln L = -\Sigma \ln x - n\ln\Gamma(b) - (1/a)\Sigma\ln x + (b-1)\Sigma\ln(\ln x) - nb\ln a \quad (2.33)$$

where the summations are taken over the n terms of the data sample.

Maximizing equation (2.33) by setting $\partial \ln L/\partial a = \partial \ln L/\partial b = 0$ gives the following maximum likelihood estimators:

$$\frac{\partial \ln L}{\partial a} = \frac{\Sigma \ln x}{a^2} - \frac{nb}{a} \quad (2.34)$$

$$\frac{\partial \ln L}{\partial b} = -n\ln a - n\psi(b) + \Sigma\ln(\ln x) = 0 \quad (2.35)$$

where the terms are as previously defined.

A comparison of equations (2.11) and (2.12) with equations (2.34) and (2.35) show that the 2PG and L2PG distributions are similar. If the data were first logarithmically transformed and then solved by equations

(2.11) and (2.12), the answers would be the same as applying (2.34) and (2.35) to untransformed data. This is analogous to the use of the normal and lognormal distributions whereby a logarithmically transformed variable is said to be normally distributed. If the moments of the logarithms of the variable satisfy equations (2.8) and (2.9), then a L2PG distribution would be in order.

The above developments are again only applicable to standard samples. Extensions to the non-standard sample follow.

2.2.1 Inclusion of Historic Information

For a sample size n of the variate x, and a further number k whose magnitudes are known to be less than the censoring threshold, x_c, the likelihood function, L, is given by equation (2.17). The cdf for the L2PG distribution such that the probability of $x<x_c$ is:

$$F(x_c) = \frac{1}{a^b x\Gamma(b)} \int_0^{x_c} e^{-(\ln x)/a} \ln x^{b-1} dx \tag{2.36}$$

The logarithmic likelihood function, lnL, is as given by equation (2.19). A change of variate is applied such that

$$y = (\ln x)/a \tag{2.37}$$

and the density function of y is a gamma variate, parameter b, and is expressed as equation (2.21). Its F(y), $F(y_c)$, and lnL can be expressed as equations (2.22), (2.23), and (2.24), respectively. From equation (2.37), $\partial y/\partial a$ is, as well, $-y/a$. Thus, equations (2.25), (2.26), and (2.27) can be applied directly for the L2PG distribution. Solution of the equations and their expressions is as described for the 2PG distribution in Section 2.1.2.

2.3 The Pearson Type III Distribution

In the late nineteenth century, Karl Pearson (1895) published a class of curves which had the ability to describe asymmetrical distributions. The third of the five curves has since been used in flood frequency analysis and is justly called the Pearson Type III (P3) distribution. It is also sometimes referred to as the three-parameter gamma distribution (Hahn and Shapiro 1967, p. 89).

The pdf of the P3 distribution is given by

$$f(x;a,b,m) = \begin{cases} \dfrac{\{\exp[-(x-m)/a]\}[(x-m)/a]^{b-1}}{|a|\Gamma(b)}, & x \geq m, a>0, b>1 \\ & \text{or } x \leq m, a<0, b>1 \\ 0 & \text{elsewhere} \end{cases} \tag{2.38}$$

where "a" and b are as previously defined and m is referred to as a location parameter. If the distribution is positively skewed, then "a" is positive and the distribution is lower bounded at m and unbounded above. Where the distribution is negatively skewed, "a" is negative and the distribution is lower bounded at zero and bounded above at m.

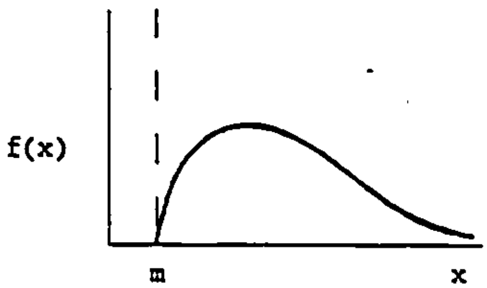
Depending on the values of the parameters "a" and b, the probability density function (pdf) described by equation (2.38) can take several shapes (Bobée 1975). However, he does not restrict b to the admissible range imposed by a maximum likelihood solution. Therefore, some of the forms that he shows are not possible with a maximum likelihood estimate of parameters and will not be shown.

A review of Figures 2.3 through 2.8 indicates that not all shapes are conducive to that commonly accepted in flood frequency analysis. Figure 2.4(a) and 2.4(b) are computationally feasible, but are not hydrologically realistic. The remaining figures show the flexibility of the distribution to assume different unimodal shapes.

Estimates of the parameters can be obtained from the moment relationships as follows (Matalas and Wallis 1973):

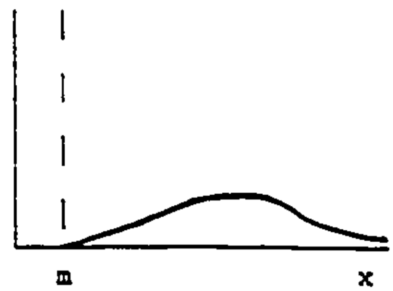
$$\bar{x} = m + ab \tag{2.39}$$
$$s = |a| \sqrt{b} \tag{2.40}$$
$$\gamma_1 = (2/\sqrt{b})(\text{sign of } a) \tag{2.41}$$
$$\gamma_2 = 3 + 6/b \tag{2.6}$$

where $\bar{x}$, s, γ_1 and γ_2 denote the mean, unbiased standard deviation, skewness, and kurtosis of the variable x. Note the similarity of equations (2.39), (2.40), and (2.41) with equations (2.3), (2.4), and (2.5) for the 2PG distribution. As parameter "a" of the P3 distribution can be either positive or negative, depending on the skewness of the sample, its absolute value is used in equation (2.40). The sign of "a" is included in equation



$1 < b < 2$

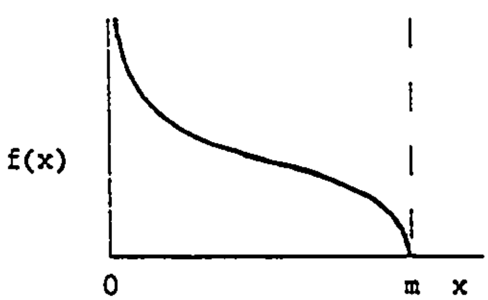
Figure 2.3(a)



$b > 2$

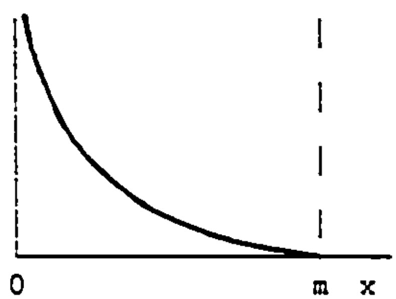
Figure 2.3(b)

Figure 2.3. The Probability Density Function of the P3 Distribution for Positive Skew, Hence Positive a



$1 < b < 2$

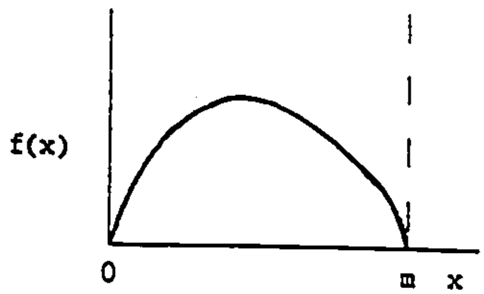
Figure 2.4(a)



$b > 2$

Figure 2.4(b)

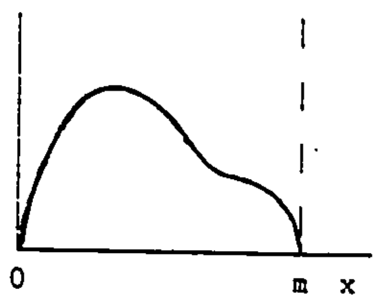
Figure 2.4. The Probability Density Function of the P3 Distribution for Negative Skew and $a < -1$



$1 < b < 2$

Figure 2.5

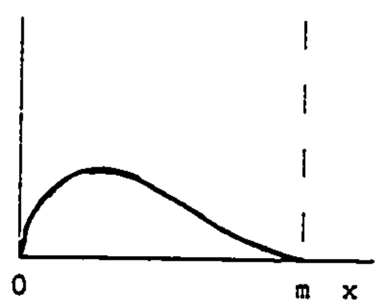
Figure 2.5. The Probability Density Function of P3 Distribution for Negative Skew and Holds if $-1<a<-1/2$ and $b \leq -(4/a^2 + 12/a+7)$ or if $-2/3<a<-1/2$ and $b>-(4/a^2 + 12/a+7)$



$1 < b < 2$

Figure 2.6

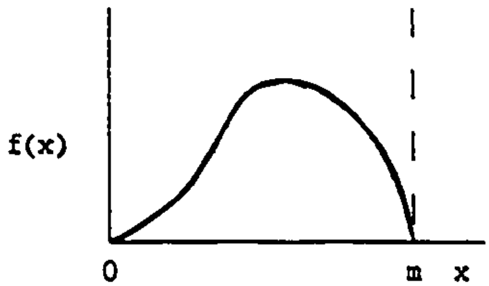
Figure 2.6. The Probability Density Function of P3 Distribution for Negative Skew and Holds if $-3/2 <a<-1$ and $b>-(4/a^2+12/a+7)$



$b > 2$

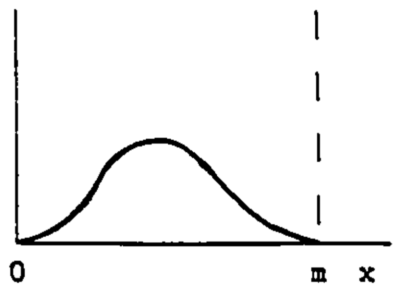
Figure 2.7

Figure 2.7. The Probability Density Function of P3 Distribution for Negative Skew and Holds if $-1<a<-1/2$ and $b>2$.



$1 < b < 2$

Figure 2.8(a)



$b > 2$

Figure 2.8(b)

Figure 2.8. The Probability Density Function of P3 Distribution for Negative Skew and $a>-1/2$

(2.41) such that the skewness will maintain its correct sign. As the square root of b is involved in equations (2.40) and (2.41), b>0. From equation (2.41), it can be seen that as b tends to infinity, the resultant skewness, γ_1, will tend to zero. The distribution's shape for large b will tend to approach normality, but the pdf will remain bounded below or above depending on the sign of "a".

The above demonstrates the complexity of the P3 distribution. It therefore can be observed that the interpretation of the results is not always trivial.

2.3.1 Maximum Likelihood Estimation of the Parameters

Fisher (1922) and Kendall and Stuart (1979, p. 70), quoting the work of Fisher, provide the logarithmic likelihood function and its first partial derivatives such that a maximum likelihood estimation of parameters can be obtained. The logarithmic likelihood function of the pdf described by equation (2.38) is:

$$\ln L = -nb\ln a - n\ln\Gamma(b) + (b-1)\Sigma\ln(x-m) - \Sigma(x-m)/a \qquad (2.42)$$

The three likelihood equations are:

$$\frac{\partial \ln L}{\partial a} = \frac{-nb}{a} + \frac{\Sigma(x-m)}{a^2} = 0 \qquad (2.43)$$

$$\frac{\partial \ln L}{\partial b} = -n\psi(b) + \Sigma\ln[(x-m)/a] = 0 \qquad (2.44)$$

$$\frac{\partial \ln L}{\partial m} = -(b-1)\Sigma(x-m)^{-1} + \frac{n}{a} = 0 \qquad (2.45)$$

where the summations are taken over the n observations and the terms are as previously defined.

Matalas and Wallis (1973) deduced from equation (2.45) that for a maximum likelihood solution of the parameters, b is constrained such that $b \geq 1$, which implies $|\gamma_1| \leq 2$. They stated that if $|\gamma_1|>2$, "the distribution is exponential" and maximum likelihood estimates from samples with $|\gamma_1|>2$ will lead to biased estimates of $|\gamma_1|$ from the parameter b.

Equations (2.43), (2.44), and (2.45) are three simultaneous transcendental equations in "a", b, and m. Equation (2.44) can be reduced to a single transcendental equation in m by rearrangement of (2.43) and (2.45) to give (Matalas and Wallis 1973):

$$a = \Sigma(x-m)/(nb) \qquad (2.46)$$

and

$$b = [n^2/\Sigma(x-m)][\Sigma(x-m)^{-1}-n^2/\Sigma(x-m)] \qquad (2.47)$$

Substitution of equation (2.46) into (2.44) eliminates "a", while substituting (2.47) into (2.44) eliminates b. Condie (1977) suggests first applying a Bolzano approach to isolate the root of equation (2.44) clear of the effect of any inflections. Once isolated, a Newton-Raphson iteration method can be used to obtain the solution. This method will yield the maximum likelihood estimates of a, b, and m for the P3 distribution.

2.3.2 Inclusion of Historic Information

As in the cases of the 2PG and L2PG distributions, the likelihood function is represented by equation (2.17). The cumulative density function, cdf, for the P3 distribution such that the probability of $x<x_c$ is:

$$F(x_c) = \frac{1}{|a|\Gamma(b)} \int_m^{x_c} \{\exp[-(x-m)/a]\}[(x-m)/a]^{b-1}dx \tag{2.48}$$

The logarithmic likelihood function, lnL, as given by equation (2.19) is applied. A change of variate is used such that

$$y = \frac{x - m}{a} \tag{2.49}$$

and the density function of y is a gamma variate, parameter b, and is expressed by equation (2.21). $F(y)$, $F(y_c)$, and lnL are as expressed in equations (2.22), (2.23), and (2.24), respectively. From equation (2.49), it is seen that $\partial y/\partial a = -y/a$. Thus equations (2.26) and (2.27) represent the first partial derivatives of the logarithmic likelihood function with respect to parameters "a" and b. The derivative of the censored term of (2.24) with respect to m is:

$$\frac{\partial \ln L}{\partial m} = \frac{k\partial}{\partial y}[\ln F(y_c)]\frac{\partial y}{\partial m} = \frac{kf(y_c)}{F(y_c)} \cdot \frac{\partial y}{\partial m} = \frac{-kf(y_c)}{aF(y_c)} \tag{2.50}$$

as $(\partial y/\partial a = -1/a)$ and the terms are as previously described. Combining equation (2.50) with equation (2.45) yields the first partial derivative of lnL with respect to m:

$$\frac{\partial \ln L}{\partial m} = \frac{-(b-1)\Sigma y^{-1}}{a} + \frac{n}{a} - \frac{kf(y_c)}{aF(y_c)} = 0 \tag{2.51}$$

Rearrangements of equations (2.26) and (2.51) yields an expression for "a":

$$a = \frac{n\,(x_c-m) - \Sigma(x-m)}{(b-1)(x_c-m)\Sigma(x-m)^{-1}-nb} \tag{2.52}$$

Substitution for "a", using equation (2.52), into equations (2.27) and (2.51) reduces the system to two simultaneous transcendental equations in b and m only. This yields two equations of the form

$$f(b,m) = 0 \tag{2.53a}$$

and

$$g(b,m) = 0 \tag{2.53b}$$

and can be solved by any numerical analysis method (Condie and Pilon 1983).

2.4 The Log Pearson Type III Distribution

If y of equation (2.49) is a gamma variate, parameter b, the distribution of x is the Pearson Type III. If x is replaced by its natural logarithm, the distribution of x becomes the Log Pearson Type III (LP3). Its hydrologic popularity is predominantly based on the decision by the Hydrology Committee of the Water Resources Council (Benson, 1968) to recommend the LP3 distribution as the standard method of analysis, with the provision that other methods could be used when it was deemed necessary.

Thomas (1985) traces the history of the application of the LP3 distribution by the U.S. Water Resources Council. Their work culminated in Bulletin 17B (Hydrology Subcommittee 1982) and represents the current guidelines used by federal agencies of the U.S.A. Possibly due to the high attention given the LP3 in hydrologic publications, the approach has been used in other countries. The LP3 is used for flood frequency analysis in Australia (Srikanthan and McMahon 1981), South Africa (Alexander 1988), and Canada (Condie et al. 1981), as well as other countries.

Most of the published literature on frequency analysis deals with one or more of the three problems Fisher (1922) felt arose in the "reduction of data". These three problems were regarding: 1) the "choice of the mathematical form of the population"; 2) the "choice of methods... to estimate the parameters of the hypothetical population"; and 3) the "distribution of statistics derived from samples". He felt that once these problems were satisfied, "then the theoretical aspect of the treatment of any particular body of data [would be] completely elucidated".

Interestingly in the same paper, he demonstrated for the P3 distribution that maximum likelihood theory provided asymptotically the most efficient estimates of parameters. This was, in particular, a response to the method of moments approach advocated by Karl Pearson, and its publication ended "one of the most colorful chapters in the history of statistical theory" (Matalas and Wallis 1973, p. 282).

Much of the hydrologic literature regarding frequency analysis can, as well, be categorized as dealing with one or more of Fisher's "problems". This is particularly evident in the case of the LP3 distribution. In the United States, dissent exists regarding its choice as the base distribution and the use of the method of moments to estimate its parameters. A review of Benson's (1968) paper indicates that the selection of the LP3 was not based on an exhaustive study. Indeed, many including Matalas and Wallis (1973) regard the adoption of the LP3 as a "choice by fiat". Others, such as, Matalas et al. (1975), Wallis and Wood (1985), and Pilon et al. (1987) deal with the "form of the population" and can be used to demonstrate the inadequacy of the LP3 as a base distribution. In addition, several papers deal with the fitting procedure by moments or are concerned with moments due to their application in the base procedures (Wallis et al. 1974; Matalas et al. 1975; Kirby 1974; Bobée and Robitaille 1975; Bobée and Robitaille 1977; Landwehr et al. 1978; Nozdryn-Plotnicki and Watt 1979), while others propose different parameter estimation techniques (Bobée 1975; Buckett and Oliver 1977; Condie 1977; Bobée 1979; Rao 1980; Hoshi and Burges 1981; Phien and Hira 1983; Singh and Singh 1985; Askar and Bobée 1987).

Bobée (1973) developed the variance of the T-year event for a sample that was fitted to a P3 distribution by moments, which subsequently can be extended to an LP3. Condie (1977) produced the asymptotic variance for the LP3 distribution based on maximum likelihood, while Phien and Hsu (1985), commented upon by Askar and Bobée (1986), were primarily concerned with the variance when the parameters are estimated by mixed moments.

The LP3 distribution represents the most widely published and applied distribution in flood frequency analysis. In view of its importance, further in-depth studies will be focused on this distribution.

2.4.1 Parameter Estimation for the Log Pearson Type III Distribution

The pdf of the LP3 distribution is given by

$$f(x;a,b,m) = \begin{cases} \dfrac{\exp[-(\ln x-m)/a][(\ln x-m)/a]^{b-1}}{|a|x\Gamma(b)}, & x \geq e^m,\ a>0,\ b>1 \\ & \text{or } o \leq x \leq e^m\ a<o,\ b>1 \\ 0 & \text{elsewhere} \end{cases} \qquad (2.54)$$

where a, b, and m are, respectively, scale, shape, and location parameters. If the distribution is positively skewed, then "a" is greater than zero and the pdf is lower bounded at e^m. Conversely, if the distribution is negatively skewed, then "a" is negative and the pdf is lower bounded at zero and upper bounded at e^m.

Bobée (1975) shows the many shapes the pdf can assume. These shapes are analgous to those shown in Figures 2.3 to 2.8, except that the boundary should be e^m rather than m.

If $y = \ln x$, estimates of the parameters can be obtained from the moment relationships by extension of the P3 equations, such that:

$$\bar{y} = m + ab \tag{2.39b}$$

$$S_y = |a|\sqrt{b} \tag{2.40b}$$

$$\gamma_{1,y} = (2/\sqrt{b})(\text{sign of } a) \tag{2.41b}$$

$$\gamma_{2,y} = 3 + 6/b \tag{2.6b}$$

where $\bar{y}$, S_y, $\gamma_{1,y}$, and $\gamma_{2,y}$ denote the mean, unbiased standard deviation, skewness, and kurtosis of the variable y. From the work of Bobée (1975), parameter b must be larger than one so that the pdf has a shape conducive to that commonly assumed in flood frequency analysis. Given a lower physical limit on b of one, equation (2.41) indicates that the skewness of a P3 or a LP3 variate cannot exceed two. Given that the annual maxima of streamflow can have skewness greater than two and that the logarithmic transformation reduces the skewness of a variate, the LP3 was preferred to the P3 for general application.

Condie (1977) provides the logarithmic likelihood function of the pdf described by equation (2.54). It is

$$\ln L = -\Sigma \ln x - n\ln\Gamma(b) - (1/a)\Sigma(\ln x-m) + (b-1)\Sigma\ln[(\ln x-m)/a] - n\ln|a| \tag{2.55}$$

where the summations are taken over the n terms of the data sample. The maximum likelihood estimators are obtained by setting $(\partial \ln L/\partial a) = (\partial \ln L/\partial b) = (\partial \ln L/\partial m) = 0$. The first partial derivatives are:

$$\frac{\partial \ln L}{\partial a} = \frac{-nb}{a} + \frac{\Sigma(\ln x-m)}{a^2} = 0 \tag{2.56}$$

$$\frac{\partial \ln L}{\partial b} = -n\psi(b) + \Sigma \ln[(\ln x-m)/a] = 0 \tag{2.57}$$

$$\frac{\partial \ln L}{\partial m} = \frac{n}{a} - (b-1)\Sigma(\ln x-m)^{-1} = 0 \tag{2.58}$$

where the terms are as previously defined.

The three first partial derivatives of equation (2.55) represent three simultaneous transcendental equations in a, b, and m. Condie (1977) reduces the system of three equations to a system containing two equations:

$$a = \frac{\Sigma(\ln x-m)}{nb} \tag{2.59}$$

$$b = \Sigma(\ln x-m)^{-1}/[\Sigma(\ln x-m)^{-1} - n^2/\Sigma(\ln x-m)] \tag{2.60}$$

These two equations can be solved using standard numerical procedures such as the Bolzano method or Newton-Raphson iteration procedure. The solution of which yields maximum likelihood estimates of the parameters.

Condie (1977) deduced that the estimate of b by this technique must exceed one if a solution is to be found. Thus, if the underlying but unknown distribution has in fact a skewness of the logarithms greater than 2 and if a maximum likelihood solution is obtained, then the estimates of the parameters will be biased. This represents an extreme skewness condition which seldom occurs in hydrometric records. However, this represents not only a constraint on the solution approach but a restraint on use of the LP3 distribution to mimic certain known shapes of a sample's pdf when the skewness of the logarithms exceeds two.

2.4.2 Inclusion of Historic Information

The likelihood function incorporating historic information is as given by equation (2.17). The cumulative density function (cdf) for the LP3 distribution such that the probability of $x<x_c$ is:

$$F(x_c) = \frac{1}{|a|\Gamma(b)} \int_{e^m}^{x_c} \frac{1}{x} \left(\frac{\ln x-m}{a}\right)^{b-1} \exp\left[-\left(\frac{\ln x-m}{a}\right)\right] dx \tag{2.61}$$

The logarithmic likelihood function defined by equation (2.19) is applied. A change of variate is used where

$$y = \frac{\ln x - m}{a} \tag{2.62}$$

and the density function of y is a gamma variate, parameter b, and is expressed by equation (2.21). $F(y)$, $F(y_c)$, and $\ln L$ are as expressed in equations (2.22), (2.23), and (2.24), respectively. The first partial derivative of equation (2.62) with respect to "a" gives $\partial y/\partial a = -y/a$. Thus, equations (2.26) and (2.27) represent the first partial derivatives of the logarithmic likelihood function with respect to parameters "a" and b. The first partial derivative with respect to m is given by equation (2.51).

Rearrangements of equations (2.26) and (2.51) gives an expression for "a":

$$a = \frac{n(\ln x_c - m) - \Sigma(\ln x - m)}{(b-1)(\ln x_c - m)\Sigma(\ln x - m)^{-1} - nb} \tag{2.63}$$

Substitution for "a", using equation (2.63), into equations (2.27) and (2.51) reduces the system to two simultaneous transcendental equations in b and m only. As in the case of the P3 distribution, this yields two equations of the form

$$f(b,m) = 0 \tag{2.53a}$$

and

$$g(b,m) = 0 \tag{2.53b}$$

and can be solved by any numerical analysis method (Condie and Pilon 1983).

2.5 The Generalized Gamma Distribution

The Generalized Gamma (GG) distribution is used extensively in Eastern Europe and the Soviet Union. In contrast, the distribution is not well known in North America. The GG distribution is also referred to as the Power Transformed Gamma distribution, the Stacy distribution, the Kritskii-Menkel distribution, and the Three-Parameter Gamma distribution (not the P3 distribution, but same name for both). Its main advantage is that the distribution remains unbounded above, even when the skewness is negative. The existence of an upper boundary can sometimes lead to the rejection of a distribution for a particular sample. This is so especially when the largest sample member is close to the theoretical upper limit of the pdf.

Stacey (1962) and Stacey and Mihram (1965) introduce and provide an indepth description of the mathematics of the distribution. They note that several commonly used distributions in science are special cases of the GG distribution. Some of these distributions include the Weibull or Gumbel III, the exponential, the 2PG, and the Rayleigh. The distribution proves to be very flexible, but it is mathematically complex to estimate parameters and their uncertainty (Lawless 1982). Therefore, its use in hydrology is not widespread.

The pdf of the GG distribution is given by

$$f(x;a,b,h) = \begin{cases} \dfrac{h(x/a)^{hb-1}\exp-(x/a)^{h}}{a\Gamma(b)}, & x\geq 0,\ a>0,\ b>0 \\ & -\infty<h<\infty \\ 0 & \text{elsewhere} \end{cases} \tag{2.64}$$

where h and b are jointly shape parameters and "a" is a scaling parameter. The pdf is bounded below at zero and is unbounded above when the skewness is both positive or negative.

Stacey and Mihram (1965) show the various shapes the pdf of the GG distribution can assume when h>0. They vary from exponential die away form and reverse J shapes to unimodal shapes commonly associated with the pdf's in flood frequency analysis. For example, when h = 1, the GG becomes the

2PG distribution. They note that the pdf has a unique smooth maximum whenever $hb<0$ or $hb>1$.

Condie (1978) derives, using the theory of moments, relationships for the parameters:

$$\bar{x} = \frac{a\Gamma(b+1/h)}{\Gamma(b)} \tag{2.65}$$

$$S = \left(\frac{a^2\Gamma(b+2/h)}{\Gamma(b)} - \frac{a^2\Gamma^2(b+1/h)}{\Gamma^2(b)} \right)^{1/2} \tag{2.66}$$

$$CV = \left(\frac{\Gamma(b+2/h)\Gamma(b)}{\Gamma^2(b+1/h)} - 1 \right)^{1/2} \tag{2.67}$$

$$\gamma_1 = \frac{1}{CV^3} \left(\frac{\Gamma^2(b)\Gamma(b+3/h)-3\Gamma(b)\Gamma(b+1/h)\Gamma(b+2/h)}{\Gamma^3(b+1/h)} + 2 \right) \tag{2.68}$$

where $\bar{x}$, S, CV, and γ_1 are the sample mean, unbiased standard deviation, coefficient of variation, and skewness of the variate x. Parameter estimates from these moment relationships could be obtained by solving the simultaneous relations of equations (2.67) and (2.68) to obtain b and h. Substitution of the estimated b and h into either equation (2.65) or (2.66) will yield "a".

Applying maximum likelihood theory to the pdf described by equation (2.64) gives the logarithmic likelihood function (Condie 1978):

$$\ln L = n\ln h - \Sigma(x/a)^h + (hb-1)\Sigma\ln x - nhb\ln a - n\ln\Gamma(b) \qquad (2.69)$$

where the summations are taken over the n terms of the data sample. The maximum likelihood estimators are obtained by setting $(\partial\ln L/\partial a) = (\partial\ln L/\partial b) = (\partial\ln L/\partial h) = 0$. This results in three simultaneous transcendental equations:

$$\frac{\partial\ln L}{\partial a} = \frac{h\Sigma x^h}{a^{h+1}} - \frac{nhb}{a} = 0 \qquad (2.70)$$

$$\frac{\partial\ln L}{\partial b} = h\Sigma\ln x - nh\ln a - n\psi(b) = 0 \qquad (2.71)$$

$$\frac{\partial\ln L}{\partial h} = \frac{n}{h} - \Sigma\left[\left(\frac{x}{a}\right)^h \ln\left(\frac{x}{a}\right)\right] + b\Sigma\ln x - nb\ln a = 0 \qquad (2.72)$$

where $\psi(b)$ represents the psi or digamma function as previously defined. Condie (1978) isolates the parameter h into one single transcendental equation which when solved leads to the parameters "a" and b. The relations are:

$$h\Sigma\ln x - n\ln(\Sigma x^h) - n\ln(nb) - n\psi(b) = 0 \qquad (2.73)$$

where

$$b = (n\Sigma x^h)/[h(n\Sigma x^h\ln x - \Sigma x^h\Sigma\ln x)] \qquad (2.74)$$

and

$$a = (\Sigma x^h/nb)^{1/h} \qquad (2.75)$$

Substitution of (2.74) into (2.73) gives the equation to be solved for h. The solution may result in more than one root which maximizes the likelihood function. Evaluation of the likelihood function with each possible parameter combination will yield the set which truly maximizes the function.

2.5.1 Inclusion of Historic Information

Equation (2.17) represents the likelihood function incorporating historic information. The cdf for the GG distribution such that the probability of $x<x_c$ is:

$$F(x_c) = \frac{h}{a\Gamma(b)} \int_0^{x_c} (x/a)^{hb-1} \exp-(x/a)^h dx \tag{2.76}$$

The logarithmic likelihood function, lnL, defined by equation (2.19) is applied. A change of variate is used where

$$y = \left(\frac{x}{a}\right)^h \tag{2.77}$$

and the density function of y is a gamma variate, parameter b, and is expressed by equation (2.21). $F(y)$, $F(y_c)$, and lnL are evaluated using equations (2.22), (2.23), and (2.24), respectively. The first partial derivative of equation (2.77) with respect to "a" gives $\partial y/\partial a = -hy/a$. The derivative of the censored term of (2.24) with respect to "a" is:

$$\frac{\partial \ln L}{\partial a} = \frac{k\partial}{\partial y} [\ln F(y_c)] \frac{\partial y}{\partial a} = \frac{kf(y_c)}{F(y_c)} \frac{\partial y}{\partial a} = \frac{-y_c hkf(y_c)}{aF(y_c)} \tag{2.78}$$

Combining equations (2.70) and (2.78) yields the first partial derivative of lnL with respect to "a":

$$\frac{\partial \ln L}{\partial a} = \frac{-nhb}{a} + \frac{h\Sigma y}{a} - \frac{y_c hkf(y_c)}{aF(y_c)} = 0 \tag{2.79a}$$

which can be reduced to:

$$\frac{\partial \ln L}{\partial a} = -nb + \Sigma y - \frac{y_c kf(y_c)}{F(y_c)} = 0 \tag{2.79b}$$

Combining equation (2.71) with the derivative of (2.24) with respect to b gives:

$$\frac{\partial \ln L}{\partial b} = \Sigma \ln y - n\psi(b) + k[\partial \ln F(y_c)/\partial b] = 0 \tag{2.80}$$

The derivative of equation (2.62) with respect to h is (ylny)/h. Thus, the derivative of the censored term of (2.24) with respect to h is:

$$\frac{\partial \ln L}{\partial h} = \frac{k\partial}{\partial y}[\ln F(y_c)]\frac{\partial y}{\partial a} = \frac{kf(y_c)}{F(y_c)} \cdot \frac{\partial y}{\partial a} = \frac{ky_c(\ln y_c)f(y_c)}{hF(y_c)} \tag{2.81}$$

Combining equations (2.72) and (2.81) gives:

$$\frac{\partial \ln L}{\partial h} = \frac{-\Sigma(y\ln y)}{h} + \frac{b\Sigma \ln y}{h} + \frac{n}{h} + \frac{ky_c(\ln y_c)f(y_c)}{hF(y_c)} = 0 \tag{2.82a}$$

which can be reduced to:

$$\frac{\partial \ln L}{\partial h} = -\Sigma(y\ln y) + b\Sigma\ln y + n + \frac{ky_c(\ln y_c)f(y_c)}{F(y_c)} = 0 \qquad (2.82b)$$

Equations (2.78), (2.79a) and (2.82a) are used in the formulation of the asymptotic error, while equations (2.78), (2.79b) and (2.82b) are used to obtain the maximum likelihood estimates of the parameters.

Combining equations (2.79b) and (2.82b) yields an expression for b:

$$b = \{\Sigma[(x/a)^h\ln(x/a)] - \ln(x_c/a)\Sigma(x/a)^h - n/h\}/$$
$$[\Sigma\ln(x/a) - n\ln(x/a)] \qquad (2.83)$$

Substitution of equation (2.83) into equations (2.78) and (2.79b) reduces the system to two simultaneous transcendental equations in "a" and h:

$$f(a,h) = 0 \qquad (2.53c)$$

$$g(a,h) = 0 \qquad (2.53d)$$

which can be solved by numerical analysis techniques.

The expression $[\partial\ln F(y_c)/\partial b]$ in equation (2.80) can be evaluated using equations (2.29b), (2.30), and (2.31) where $y_c = (x_c/a)^h$ and x_c is the censoring threshold.

In summary, this chapter contains the appropriate theoretical developments necessary for parameter estimation. These developments will now be used in the next chapter to derive the asymptotic standard error of the parameter and quantile estimates.

CHAPTER 3

THEORETICAL DEVELOPMENTS OF THE ASYMPTOTIC STANDARD ERROR OF ESTIMATE OF PARAMETERS AND QUANTILES

3.1 Introduction

Fisher (1922) presented a method by which the asymptotic estimates of parameter variances and covariances could be obtained from the logarithmic likelihood function. In this section, Fisher's (1922) method will be given in general form. Subsequent sections of this chapter present applications of this procedure to the gamma-type distributions of Chapter 2.

If the probability density function (pdf) of a random variable is given as $f(x;\alpha_1,\alpha_2,\ldots\alpha_p)$, then the logarithmic likelihood function, lnL, is:

$$\ln L = \Sigma \ln\,[f(x;\alpha_1,\alpha_2,\ldots\alpha_p)] \tag{3.1}$$

where $\alpha_1,\alpha_2,\ldots\alpha_p$ represent the p parameters of the distribution and the summation is taken over all values of x. First-order and second-order partial derivatives of lnL with respect to the parameters can be found. The elements of the inverse variance-covariance matrix or the inverse dispersion matrix are given by:

$$r_{i,j} = \begin{cases} -E(\partial^2 \ln L/\partial\alpha_i \ \partial\alpha_j) & , \quad i \neq j \\ -E(\partial^2 \ln L/\partial\alpha_i{}^2) & , \quad i = j \end{cases} \tag{3.2}$$

where E is the expected value of the argument within parentheses. Inversion of the inverse dispersion matrix yields the variances and covariances of the parameters:

$$[r_{i,i}]^{-1} = \mathrm{var}(\alpha_i)$$

and

$$[r_{i,j}]^{-1} = \mathrm{covar}(\alpha_i,\alpha_j)$$

Having obtained the asymptotic variances and covariances of the parameters, the general relationship for the variance of a function of p statistical variables can be used to obtain the asymptotic variance of estimate of the quantiles. That is,

$$\begin{aligned} \mathrm{var}(x) = {} & (\partial x/\partial\alpha_1)^2 \ \mathrm{var}(\alpha_1) + (\partial x/\partial\alpha_2)^2 \ \mathrm{var}(\alpha_2) \\ & + \dots (\partial x/\partial\alpha_p)^2 \ \mathrm{var}(\alpha_p) + 2(\partial x/\partial\alpha_1)(\partial x/\partial\alpha_2) \\ & \mathrm{covar}(\alpha_1,\alpha_2) + \dots 2(\partial x/\partial\alpha_1)(\partial x/\partial\alpha_p) \ \mathrm{covar}(\alpha_1,\alpha_p) \\ & + 2(\partial x/\partial\alpha_2)(\partial x/\partial\alpha_3) \ \mathrm{covar}(\alpha_2,\alpha_3) + \dots 2(\partial x/\partial\alpha_2) \\ & (\partial x/\partial\alpha_p) \ \mathrm{covar}(\alpha_2,\alpha_p) + \dots \end{aligned} \tag{3.3}$$

where x is the quantile or T-year flood and var(x) is the asymptotic variance of estimate of the T-year event. The square root of var(x) gives the asymptotic standard error of the estimate.

The above-outlined theory will now be implemented for the distributions used in flood frequency analysis.

3.2 The Two Parameter Gamma Distribution

The pdf of the two parameter gamma (2PG) distribution is given by equation (2.1) and its logarithmic likelihood function is given by equation (2.10). The second partial derivatives of lnL with respect to the parameters of the 2PG are:

$$\partial^2 \ln L/\partial a^2 = nb/a^2 - 2(\Sigma x)/a^3 \tag{3.4}$$

$$\frac{\partial^2 \ln L}{\partial b^2} = \frac{-n\partial^2 \ln\Gamma(b)}{\partial b^2} = -n\psi'(b) \tag{3.5}$$

$$\frac{\partial^2 \ln L}{\partial a \partial b} = \frac{-n}{a} \tag{3.6}$$

where $\psi'(b)$ is the trigamma function. As $E(\Sigma x)$ for the 2PG is nab, the negative expectation of (3.4) is

$$-E(\partial^2 \ln L/\partial a^2) = nb/a^2 = r_{1,1} \tag{3.7}$$

The negative expectation of (3.5) and (3.6) are

$$-E(\partial^2 \ln L/\partial b^2) = n\psi'(b) = r_{2,2} \tag{3.8}$$

and

$$-E(\partial^2 \ln L/\partial a \partial b) = n/a = r_{1,2} = r_{2,1} \tag{3.9}$$

Thus, the inverse dispersion matrix is then

$$V^{-1} = n \begin{vmatrix} b/a^2 & 1/a \\ 1/a & \psi'(b) \end{vmatrix}$$

Inversion of the above matrix yields

$$\mathrm{var}(a) = \frac{a^2 \ \psi'(b)}{n[b\psi'(b)-1]} \tag{3.10}$$

$$\mathrm{var}(b) = \frac{b}{n[b\psi'(b)-1]} \tag{3.11}$$

$$\mathrm{cov}(a,b) = \frac{a}{n[b\psi'(b)-1]} \tag{3.12}$$

By use of the Wilson-Hilferty transform, the T-year event for the 2PG distribution can be obtained:

$$x_T = a[t/(3b^{1/6}) - 1/(9b^{2/3}) + b^{1/3}]^3 \tag{3.13}$$

where a and b are as previously defined and t is the standard normal deviate. From equation (3.13), the following first-order partial derivates can be derived:

$$\frac{\partial x}{\partial a} = [t/(3b^{1/6}) - 1/(9b^{2/3}) + b^{1/3}]^3 \tag{3.14}$$

$$\frac{\partial x}{\partial b} = 3a[t/(3b^{1/6}) - 1/(9b^{2/3}) + b^{1/3}]^2\ [-t/(18b^{7/6}) + 2/(27b^{5/3}) + 1/(3b^{2/3})] \tag{3.15}$$

The equation for the asymptotic variance of the T-year event can now be evaluated as an expression of a function of two variables, each subject to sampling variance and covariance:

$$\mathrm{var}(x_T) = (\partial x/\partial a)^2\mathrm{var}(a) + (\partial x/\partial b)^2\mathrm{var}(b) + 2(\partial x/\partial a)(\partial x/\partial b)\mathrm{cov}(a,b) \tag{3.3a}$$

where $(\partial x/\partial a)$, $(\partial x/\partial b)$, var(a), var(b), and cov(a,b) are as estimated by equations (3.14), (3.15), (3.10), (3.11), and (3.12), respectively.

3.2.1 Type-I Censoring

The logarithmic likelihood function when dealing with the censored sample is represented by equation (2.19). Equations (2.26) and (2.27) represent the first-order partial derivatives of (2.19) for the 2PG distribution. The second-order partial derivatives can be found using the change of variate technique of equation (2.20). They are:

$$\frac{\partial^2 \ln L}{\partial a^2} = \frac{-2\Sigma y}{a^2} + \frac{nb}{a^2} - \frac{ky_c f(y_c)}{aF(y_c)}\left[\frac{-2}{a} - \frac{y_c f'(y_c)}{af(y_c)} + \frac{y_c f(y_c)}{aF(y_c)}\right] \tag{3.16}$$

$$\frac{\partial^2 \ln L}{\partial b^2} = -n\ \psi'(b) + \frac{k[\partial F(y_c)/\partial b]}{F(y_c)}\left[\frac{\partial \ln[\partial F(y_c)/\partial b]}{\partial b} - \frac{(\partial F(y_c)/\partial b)}{F(y_c)}\right] \tag{3.17}$$

$$\frac{\partial^2 \ln L}{\partial a \partial b} = \frac{-n}{a} - \frac{k y_c f(y_c)}{a F(y_c)} \left[\ln y_c - \psi(b) - \frac{f(t_c)}{F(t_c)} \frac{\partial t_c}{\partial b} \right] \tag{3.18}$$

where $f(y_c)$, $F(y_c)$, and t_c are found using equations (2.21), (2.23), and (2.29b): $\psi^1(b)$ is the trigamma function $[\partial\psi(b)/\partial b]$; and $f^1(y_c)$, $f(t_c)$, $[\partial F(y_c)/\partial b]$, and $\{\partial[\ln(\partial F(y_c)/\partial b)]/\partial b\}$ are found from:

$$f^1(y_c) = f(y_c) \left[\frac{(b-1)}{y_c} - 1 \right] \tag{3.19}$$

$$f(t_c) = \frac{e^{-t_c^{2/2}}}{\sqrt{2\pi}} \tag{3.20}$$

$$\frac{\partial F(y_c)}{\partial b} = f(t_c) \frac{\partial t_c}{\partial b} \tag{3.21}$$

$$\frac{\partial[\ln(\partial F(y_c)/\partial b)]}{\partial b} = \left(\frac{\partial^2 t_c}{\partial b^2} \Big/ \frac{\partial t_c}{\partial b} \right) - t_c \frac{\partial t_c}{\partial b} \tag{3.22}$$

where

$$\frac{\partial^2 t_c}{\partial b^2} = \frac{-5 y_c^{1/3}}{12 b^{11/6}} + \frac{1}{4 b^{5/2}} + \frac{3}{4 b^{3/2}} \tag{3.23}$$

When a is negative, as sometimes occurs in the Pearson Type III and log Pearson Type III distributions, then the sign of $f^1(y_c)$, t_c, $(\partial t_c/\partial b)$, and $\partial^2 t_c/\partial b^2$ would be changed. $f(t_c)$ and $F(t_c)$ represent the ordinate and the area under the normal curve, respectively. Note that $F(y_c)$ is equivalent in value to $F(t_c)$, as both represent the probability of $x \leq x_c$ or $y \leq y_c$.

The expectation of the (Σy) term in equation (3.16) is derived from the asymptotic form of the Type-I censored sample. Figure 3.1 shows a partially truncated distribution which represents the Type-I censored sample case.

n_a is the number of fully defined floods $\geq x_c$, n_b is the number of fully defined floods $< x_c$, n_c or k is the number of censored floods $< x_c$, and YT is the historic time span of the analysis and is equal to n + k, where n and k are as previously defined - the number of fully defined floods and the number of censored floods $< x_c$. For Type-I censoring, Condie (1986) notes that (n_a/YT) will tend to $(1-F(x_c))$ and $[n_c/(n_b+n_c)]$ will tend to q, a constant ratio. The hatched portion of the pdf of Figure 3.1 represents the partially truncated distribution. It is defined by:

$$(1-q)f(x) \quad \text{for} \quad x < x_c$$

and

$$f(x) \quad \text{for} \quad x > x_c.$$

Its area is no longer equal to unity, thus it is necessary to divide the partially truncated distribution by its area, $[1-qF(x_c)]$.

The expected value of Σy is comprised of two integrals, one from the lower boundary to x_c in the partially truncated region, while the second is from x_c to the upper bound. The expectation is defined by the integral

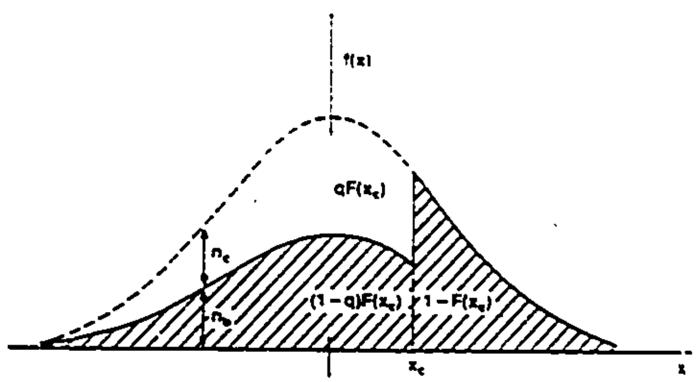


Figure 3.1. The Partially Truncated Distribution (Condie 1986, p. 149)

$$\int \frac{y\, e^{-y}\, y^{b-1}}{\Gamma(b)}\, dy$$

or

$$\frac{\Gamma(b+1)}{\Gamma(b)} \int \frac{e^{-y}\, y^{b}}{\Gamma(b+1)}\, dy$$

Thus, the total integral is

$$b\ \{(1-q)F[y_c,(b+1)] + 1 - F[y_c,(b+1)]\}$$

where $F[y_c,(b+1)]$ is the incomplete gamma function with parameter (b-1). Previously, $F(y_c)$ was defined as the incomplete gamma function with parameter b, the addition of the parameter within the square-bracketed term is used to differentiate the various incomplete gamma functions required in the evaluation of the expectations.

However, the total integral must be divided by the area under the partially truncated distribution. Thus,

$$E(y) = \frac{b\ \{1-qF[y_c,(b+1)]\}}{\{1-qF[y_c,(b)]\}} \tag{3.24}$$

where $q = n_c/(n_b+n_c)$ and b is a parameter as previously defined. Equation (3.24) is evaluated and is subsequently substituted into equation (3.16).

The expectations of equations (3.16), (3.17), and (3.18) can now be estimated. Insertion of these values into the inverse dispersion matrix and its subsequent inversion provides the variances and covariances of the parameters estimated from a Type-I censored sample. The asymptotic variance of a quantile estimate can be obtained by substitution of the inverted matrix elements and $(\partial x/\partial a)$ and $(\partial x/\partial b)$ from equations (3.14) and (3.15) into equation (3.3a).

3.3 The Log Two Parameter Gamma Distribution

The pdf of the L2PG distribution is given by equation (2.32). Equation (2.33) is its logarithmic likelihood function, while equations (2.34) and (2.35) are the first-order partial derivatives of lnL with respect to the parameters of the distribution. The second-order partial derivatives of lnL are:

$$\partial^2 \ln L/\partial a^2 = nb/a^2 - 2(\Sigma \ln x)/a^3 \tag{3.25}$$

$$\partial^2 \ln L/\partial b^2 = -n\psi'(b) \tag{3.5}$$

$$\partial^2 \ln L/\partial a \partial b = -n/a \tag{3.6}$$

The $E(\Sigma \ln x)$ is nab for the L2PG distribution. Thus the elements of the inverse dispersion matrix are as given for the 2PG distribution. That is, equations (3.7), (3.8), and (3.9) can be used to fill the matrix, while equations (3.10), (3.11), and (3.12) represent elements of the inverted matrix and are the asymptotic estimates of the variances and covariances of the parameters.

The T-year event for the L2PG distribution can be obtained by use of the Wilson-Hilferty transform as:

$$\ln x_T = a[t/(3b^{1/6} - 1/(9b^{2/3}) + b^{1/3}]^3 \qquad (3.13a)$$

where the terms are as previously defined. If a change of variate is applied such that:

$$Z_T = \ln x_T \qquad (3.26)$$

Then, the asymptotic variance is

$$\mathrm{var}(Z_T) = (\partial z/\partial a)^2 \mathrm{var}(a) + (\partial z/\partial b)^2 \mathrm{var}(b) + 2(\partial z/\partial a)(\partial z/\partial b)\mathrm{cov}(a,b) \qquad (3.3b)$$

where $(\partial z/\partial a)$, $(\partial z/\partial b)$, var(a), var(b), and cov(a,b) are as estimated by equations (3.14), (3.15), (3.10), (3.11), and (3.12), respectively. Note that $(\partial z/\partial a) = (\partial x/\partial a)$ and $(\partial z/\partial b) = (\partial x/\partial b)$ in equations (3.14) and (3.15), respectively.

As a change of variate has been used, the variance of the untransformed quantile, x_T, can be found from:

$$\mathrm{var}(x_T) = (dx/dz)^2\ \mathrm{var}(\ln x_T) \qquad (3.27)$$

where $[(dx/dz) = x]$. Hence, the standard error of estimate of the T-year event for the L2PG variate is:

$$\text{S.E. of } x_T = x_T \, [\mathrm{var}(\ln x_T)]^{1/2} \tag{3.28}$$

3.3.1 Type-I Censoring

The logarithmic likelihood function for the censored sample case is as given in equation (2.19). A change of variate as expressed in equation (2.37) is applied and the resultant first-order partial derivatives of the lnL with respect to the parameters are as given by equations (2.26) and (2.27). The second-order partials are as those found for the 2PG distribution and are given in equations (3.16), (3.17), and (3.18). Their expectations, as described in Section 3.2.2, form the elements of inverse-dispersion matrix. Inversion of this matrix results in the estimate of the asymptotic variances and covariances of the parameters from a Type-I censored sample. The asymptotic variance of a quantile can be obtained through the use of the change of variate of equation (3.26) and equations (2.13a), (3.3b), and (3.27).

3.4 The Pearson Type III Distribution

The pdf of the P3 distribution is represented by equation (2.38). Its logarithmic likelihood equation is given in equation (2.42), while the first-order partial derivatives with respect to the parameters are given in equations (2.43), (2.44), and (2.45). From the work of Fisher (1922), which is also quoted and possibly more readily available in Kendall and Stuart (1976), the second-order partial derivatives are:

$$\partial^2 \ln L/\partial a^2 = -(2/a^3)\Sigma(x-m) + nb/a^2 \qquad (3.29)$$

$$\partial^2 \ln L/\partial b^2 = -n\psi^1(b) \qquad (3.5)$$

$$\partial^2 \ln L/\partial m^2 = -(b-1)\Sigma(x-m)^{-2} \qquad (3.30)$$

$$\partial^2 \ln L/\partial a\partial b = -n/a \qquad (3.6)$$

$$\partial^2 \ln L/\partial a\partial m = -n/a^2 \qquad (3.31)$$

$$\partial^2 \ln L/\partial b\partial m = -\Sigma(x-m)^{-1} \qquad (3.32)$$

where the terms are as previously defined. The elements of the inverse-dispersion matrix are the negative expectations of the equations of the second-order partial derivatives. The elements are:

$$r_{1,1} = -E(\partial^2 \ln L/\partial a^2) = nb/a^2 \tag{3.7}$$

$$r_{2,2} = -E(\partial^2 \ln L/\partial b^2) = n\psi^1(b) \tag{3.8}$$

$$r_{3,3} = -E(\partial^2 \ln L/\partial m^2) = n/[a^2(b-2)] \tag{3.33}$$

$$r_{1,2} = r_{2,1} = -E(\partial^2 \ln L/\partial a \partial b) = n/a \tag{3.9}$$

$$r_{1,3} = r_{3,1} = -E(\partial^2 \ln L/\partial a \partial m) = n/a^2 \tag{3.34}$$

$$r_{2,3} = r_{3,2} = -E(\partial^2 \ln L/\partial b \partial m) = n/[a(b-1)] \tag{3.35}$$

The inverse dispersion matrix is then

$$V^{-1} = n \left| \begin{array}{ccc} b/a^2 & 1/a & 1/a^2 \\ 1/a & \psi^1(b) & a/[a(b-1)] \\ 1/a^2 & 1/[a(b-1)] & 1/[a^2(b-2)] \end{array} \right|$$

Inverting the matrix V^{-1} gives the estimated sampling variances and covariances. Condie (1977), working on the LP3 distribution, provides the appropriate formulation for the P3 distribution as well. The determinant D of V^{-1} is:

$$D = n^3\{[2\psi^1(b)-2/(b-1)+1/(b-1)^2]/[(b-2)a^4]\} \tag{3.36}$$

The variances and covariances of the parameters can then be expressed as:

$$var(a) = n^2[\psi^1(b)/(b-2)-1/(b-1)^2]/(Da^2) \tag{3.37}$$

$$var(b) = 2/[nDa^4(b-2)] \tag{3.38}$$

$$var(m) = [b\psi^1(b)-1]/(nDa^2) \tag{3.39}$$

$$cov(a,b) = -1[1/(b-2)-1/(b-1)]/(nDa^3) \tag{3.40}$$

$$cov(a,m) = [1/(b-1)-\psi^1(b)]/(nDa^2) \tag{3.41}$$

$$cov(b,m) = -[b/(b-1)-1]/(nDa^3) \tag{3.42}$$

Use is again made of the Wilson-Hilferty transformation such that the T-year event for the P3 distribution can be estimated from:

$$x_T = m+a\ [t/(3b^{1/6})-1/(9b^{2/3})+b^{1/3}]^3 \tag{3.43}$$

where the terms are as previously defined. The first-order partial derivatives of equation (3.43) with respect to "a" and b are given by equations (3.14) and (3.15), respectively. The first-order partial of (3.43) with respect to m is:

$$\partial x/\partial m = 1 \tag{3.44}$$

The asymptotic variance of a T-year event can be estimated as an expression of a function of three variables, each subject to sampling variance and covariance. The expression is from the general form of equation (3.3) and is:

$$\begin{aligned} \mathrm{var}(x_T) = {} & (\partial x/\partial a)^2 \mathrm{var}(a) + (\partial x/\partial b)^2 \mathrm{var}(b) + (\partial x/\partial m)^2 \mathrm{var}(m) \\ & + 2(\partial x/\partial a)(\partial x/\partial b)\mathrm{cov}(a,b) + 2(\partial x/\partial a)(\partial x/\partial m) \\ & \mathrm{cov}(a,m) + 2(\partial x/\partial b)(\partial x/\partial m)\mathrm{cov}(b,m) \end{aligned} \tag{3.3c}$$

where $(\partial x/\partial a)$, $(\partial x/\partial b)$, $(\partial x/\partial m)$, var(a), var(b), var(m), cov(a,b), cov(a,m), and cov(b,m) are estimated from equations (3.14), (3.15), (3.44) and equations (3.37) through (3.42), respectively.

3.4.1 Type-I Censoring

The logarithmic likelihood function for the censored sample case is, as well, represented by equation (2.19). Equations (2.26), (2.27), and (2.51) represent the first-order partial derivatives of the lnL function with respect to the parameters of the P3 distribution when the change of variate of equation (2.20) is used.

The second-order partial derivatives of the lnL function can be found using the change of variate of (2.20). Equations (3.16), (3.17), and (3.18) represent the $(\partial^2 \ln L/\partial a^2)$, $(\partial^2 \ln L/\partial b^2)$, and $(\partial^2 \ln L/\partial a \partial b)$ derivatives, respectively. The remaining second-order partial derivatives are:

$$\partial^2 \ln L/\partial m^2 = \frac{k}{a^2}\left[\frac{f'(y_c)}{F(y_c)} - \left(\frac{f(y_c)}{F(y_c)}\right)\right]^2 - \frac{(b-1)}{a^2}\ \Sigma(y^{-2}) \qquad (3.45)$$

$$\partial^2 \ln L/\partial a\partial m = \frac{-n}{a^2} + \frac{k}{a^2}\left[\frac{y_c f'(y_c)}{F(y_c)} - y_c\left(\frac{f(y_c)}{F(y_c)}\right)^2 + \frac{f(y_c)}{F(y_c)}\right] \qquad (3.46)$$

$$\partial^2 \ln L/\partial b\partial m = \frac{-\Sigma(y^{-1})}{a}\quad \frac{kf(y_c)}{aF(y_c)}\left[\ln y_c - \psi(b) - \frac{\partial[\ln F(y_c)]}{\partial b}\right] \qquad (3.47)$$

where

$$\frac{\partial[\ln F(y_c)]}{\partial b} = \frac{f(t_c)}{F(t_c)}\,\frac{\partial t_c}{\partial b} \qquad (3.48)$$

and the terms are as previously defined. Note that for negatively skewed samples, parameter "a" is negative. Correspondingly, the sign of the terms $f'(y_c)$, t_c, $(\partial t_c/\partial b)$, and $(\partial^2 t_c/\partial b^2)$ will be altered.

The expected value for the (y) term in the censored sample is used in equation (3.16) and is given in equation (3.24). Equations (3.45) and (3.47) contain the expressions (y^{-2}) and (y^{-1}), respectively; and, hence, expectations for these must be derived. Following the derivation of the E(y) in Section 3.2.2, the expectation of (y^{-1}) is defined by the integral

$$\int \frac{y^{-1}\, e^{-y}\, y^{b-1}}{\Gamma(b)}\, dy$$

or

$$\frac{\Gamma(b-1)}{\Gamma(b)} \int \frac{e^{-y}\, y^{b-2}}{\Gamma(b-1)}\, dy = \frac{1}{(b-1)}\, F[y_c,(b-1)]$$

where $F[y_c,(b-1)]$ is the incomplete gamma function with parameter $(b-1)$. Thus, the expectation of y^{-1} for the partially truncated density function of the P3 is:

$$E(y^{-1}) = \frac{1}{(b-1)} \{1-qF[y_c,(b-1)]\}/\{1-qF[y_c,(b)]\} \tag{3.49}$$

where the terms are as previously defined.

The expectation of y^{-2} for the partially truncated density function of the P3 distribution, as represented in Figure 3.1, is defined by the integral

$$\int \frac{y^{-2}\, e^{-y}\, y^{b-1}}{\Gamma(b)}\, dy$$

or

$$\frac{\Gamma(b-2)}{\Gamma(b)} \int \frac{e^{-y}\, y^{b-3}}{\Gamma(b-2)}\, dy = \frac{1}{(b-1)(b-2)}\, F[y_c,(b-2)]$$

Thus, the expectation of the y^{-2} term correcting for unit area is

$$E(y^{-2}) = \frac{1}{(b-1)(b-2)} \{1-qF[y_c,(b-2)]\}/\{1-qF[y_c,(b)]\} \tag{3.50}$$

Substitution of equation (3.50) into (3.45), and (3.49) into (3.47), and multiplying by n as it is a summation, results in the asymptotic approximation of expectation of $(\partial^2 \ln L/\partial m^2)$ and $(\partial^2 \ln L/\partial b \partial m)$. Substitution of equation (3.23) into (3.16), multiplied by n, yields the expectation of $(\partial^2 \ln L/\partial a^2)$. Taking the opposite sign of these expectations yields three of the six required negative expectations for the inverse-dispersion matrix. Direct evaluation of equations (3.17), (3.18), and (3.46) accompanied with a change of sign completes the required negative expectations. The inverse-dispersion matrix is inverted to obtain the asymptotic variances and covariances for parameters of the P3 distributed fitted to a Type-I censored sample. The asymptotic variance of the T-year event is obtained using equation (3.3c).

3.5 The Log Pearson Type III Distribution

The pdf of the LP3 distribution is given by equation (2.54). The lnL of this distribution is given by equation (2.55), while the first-order partial derivatives with respect to the parameters are given by equations (2.56), (2.57), and (2.58). Condie (1977) gives the second-order partial derivatives as;

$$\partial^2 \ln L/\partial a^2 = -(2/a^3)\Sigma(\ln x-m) + nb/a^2 \tag{3.51}$$

$$\partial^2 \ln L/\partial b^2 = -n\psi^1(b) \tag{3.5}$$

$$\partial^2 \ln L/\partial m^2 = -(b-1)\Sigma(\ln x-m)^{-2} \tag{3.52}$$

$$\partial^2 \ln L/\partial a\partial b = -n/a \tag{3.6}$$

$$\partial^2 \ln L/\partial a\partial m = -n/a^2 \tag{3.31}$$

$$\partial^2 \ln L/\partial b \partial m = -\Sigma(\ln x-m)^{-1} \tag{3.53}$$

where the terms are as previously defined.

The elements of the inverse-dispersion matrix are obtained from the negative expectations of the second-order partial derivatives of the lnL and are:

$$r_{1,1} = -E(\partial^2 \ln L/\partial a^2) = nb/a^2 \tag{3.7}$$

$$r_{2,2} = -E(\partial^2 \ln L/\partial b^2) = n\psi^1(b) \tag{3.8}$$

$$r_{3,3} = -E(\partial^2 \ln L/\partial m^2) = n/[a^2(b-2)] \tag{3.33}$$

$$r_{1,2} = -E(\partial^2 \ln L/\partial a \partial b) = n/a \tag{3.9}$$

$$r_{1,3} = -E(\partial^2 \ln L/\partial a \partial m) = n/a^2 \tag{3.34}$$

$$r_{2,3} = -E(\partial^2 \ln L/\partial b \partial m) = n/[a(b-1)] \tag{3.35}$$

There are identical to those obtained for the P3 distribution; thus, the inverse-dispersion matrix and its subsequent inversion will also be identical.

The T-year events for the LP3 variate can be estimated through the use of the Wilson-Hilferty transformation as:

$$\ln x_T = m + a\,[t/(3b^{1/6}) - 1/(9b^{2/3}) + b^{1/3}]^3 \qquad (3.43a)$$

where the terms are as previously defined. If a change of variate of equation (3.26) is applied, then the asymptotic variance of the new variate, Z_T, is:

$$\begin{aligned}\mathrm{var}(Z_T) = {} & (\partial z/\partial a)^2 \mathrm{var}(a) + (\partial z/\partial b)^2\, \mathrm{var}(b) + (\partial z/\partial m)^2 \\ & \mathrm{var}(m) + 2(\partial z/\partial a)(\partial z/\partial b)\mathrm{cov}(a,b) + \\ & 2(\partial z/\partial a)(\partial z/\partial m)\mathrm{cov}(a,m) + 2(\partial z/\partial b)(\partial z/\partial m) \\ & \mathrm{cov}(b,m) \qquad (3.3d)\end{aligned}$$

where $(\partial z/\partial a)$, $(\partial z/\partial b)$, $(\partial a/\partial m)$, var(a), var(b), var(m), cov(a,b), cov(a,m), and cov(b,m) are estimated by equations (3.14), (3.15), (3.44), (3.37), (3.38), (3.39), (3.40), (3.41), and (3.42), respectively. Note that $(\partial z/\partial a) = (\partial x/\partial a)$, $(\partial z/\partial b) = (\partial x/\partial b)$, and $(\partial z/\partial m) = (\partial x/\partial m)$ in equations (3.14), (3.15), and (3.44), respectively.

The asymptotic variance of the untransformed variate, x_T, is obtained by use of equation (3.27). The standard error of estimate of the untransformed variate is expressed by equation (3.28).

3.5.1 Type-I Censoring

The first-order partial derivatives of the lnL of equation (2.19) for the LP3 distribution are represented by equations (2.26), (2.27), and (2.51). A change of variate as defined by equation (2.62) is applied in the censored case. The second-order partial derivatives and the required expectations are identical to those derived for the P3 distribution. Once the asymptotic variances and covariances of the parameters have been obtained, the asymptotic variance of the untransformed variate, x_T, is obtained by use of equation (3.27). The variance of the transformed variate, used in equation (3.27), is obtained from equation (3.3d).

3.6 The Generalized Gamma Distribution

The pdf of the GG distribution is defined by equation (2.64). Its logarithmic likelihood function is given in equation (2.69). The first-order partial derivatives of equation (2.69) with respect to the parameters of the distribution are given in equations (2.70), (2.71), and (2.72). The second-order partial derivates of equation (2.69) with respect to the parameters are:

$$\partial^2 \ln L/\partial a^2 = -h(h+1)[\Sigma(x/a)^h]/a^2 \tag{3.54}$$

$$\partial^2 \ln L/\partial b^2 = -n\psi^1(b) \tag{3.5}$$

$$\partial^2 \ln L/\partial h^2 = -n/h^2-\Sigma[(x/a)^h \ln(x/a)\ln(x/a)] \tag{3.55}$$

$$\partial^2 \ln L/\partial a\partial b = -nh/a \tag{3.56}$$

$$\partial^2 \ln L/\partial a\partial h = \{\Sigma[(x/a)^h \ln(x/a)]\}/a + [\Sigma(x/a)^h]/a - nb/a \tag{3.57}$$

$$\partial^2 \ln L/\partial b\partial h = \Sigma \ln x - n\ln a \tag{3.58}$$

The negative expectations or the elements of the inverse dispersion matrix are:

$$r_{1,1} = -E(\partial^2 \ln L/\partial a^2) = h(h+1)nb/a \tag{3.59}$$

$$r_{2,2} = -E(\partial^2 \ln L/\partial b^2) = n\psi(b) \tag{3.8}$$

$$r_{3,3} = -E(\partial^2 \ln L/\partial h^2) \tag{3.60}$$

$$r_{1,2} = -E(\partial^2 \ln L/\partial a \partial b) = nh/a \tag{3.61}$$

$$r_{1,3} = -E(\partial^2 \ln L/\partial a \partial h) \tag{3.62}$$

$$r_{2,3} = -E(\partial^2 \ln L/\partial b \partial h) \tag{3.63}$$

Elements $r_{3,3}$, $r_{1,3}$, and $r_{2,3}$ must be estimated directly from the sample as their expectations have not been derived. Once the elements have been estimated, the matrix is inverted to provide the variances and covariances of the parameters of the distribution.

Any T-year event, x_T, can be estimated for the GG distribution by application of the Wilson-Hilferty transformation, whereby:

$$x_T = a[t/(3b^{1/6}) - 1/(9b^{2/3}) + b^{1/3}]^{3/h} \tag{3.64}$$

From equation (3.64), the following first-order partial derivatives may be acquired:

$$\partial x/\partial a = [t/(3b^{1/6})-1/(9b^{2/3}) + b^{1/3}]^{3/h} \tag{3.65}$$

$$\partial x/\partial b = (3a/h)[t/(3b^{1/6})-1/(9b^{2/3})+b^{1/3}]^{(3/h)-1} \cdot [-t/(18b^{7/6})+2/(27b^{5/3})+1/(3b^{2/3})] \tag{3.66}$$

$$\partial x/\partial h = -3h^{-2}a[t/(3b^{1/6})-1(9b^{2/3})+b^{1/3}]^{3/h} \cdot \ln[t/(3b^{1/6})-1/(9b^{2/3})+b^{1/3}] \tag{3.67}$$

The asymptotic variance of the T-year event can be obtained from the expression:

$$\begin{aligned} var(x_T) = {} & (\partial x/\partial a)^2 var(a) + (\partial x/\partial b)^2 var(b) + (\partial x/\partial h)^2 var(h) \\ & + 2(\partial x/\partial a)(\partial x/\partial b)cov(a,b) + 2(\partial x/\partial a)(\partial x/\partial h)cov(a,h) \\ & + 2(\partial x/\partial b)(\partial x/\partial h)cov(b,h) \end{aligned} \tag{3.3e}$$

where the terms are as previously defined.

3.6.1 Type-I Censoring

The first-order partial derivatives of the lnL for the GG distribution for the censored sample case are given in equations (2.78), (2.79a), and (2.82a). Using the change of variate defined by equation (2.77), the second-order partial derivatives of the lnL for the censored case are found to be:

$$\partial^2 \ln L/\partial a^2 = \frac{h(h+1)}{a^2}\Sigma y + \frac{nhb}{a^2} - \frac{khy_c f(y_c)}{aF(y_c)}\left\{\frac{-h}{a} - \frac{hy_c}{a}\left[\frac{(b-1)}{y_c} - 1\right] + \frac{hy_c f(y_c)}{aF(y_c)} - \frac{1}{a}\right\} \tag{3.68}$$

$$\partial^2 \ln L/\partial b^2 = -n\psi'(b) - k\partial^2[\ln F(y_c)]/\partial b^2 \tag{3.69}$$

$$\partial^2 \ln L/\partial h^2 = -\frac{n}{h^2} - \frac{1}{h^2}\Sigma[y(\ln y)^2] + \frac{ky_c^{\,2}(\ln y_c)^2}{hF(y_c)}\left\{\frac{1}{y_c} + \frac{f'(y_c)}{f(y_c)} - \frac{f(y_c)}{F(y_c)}\right\} \tag{3.70}$$

$$\partial^2 \ln L/\partial a\partial b = \frac{-nh}{a} - \frac{khy_c f(y_c)}{aF(y_c)}\left\{\ln y_c - \psi(b) - \frac{\partial[\ln F(y_c)]}{\partial b}\right\} \tag{3.71}$$

$$\partial^2 \ln L/\partial a\partial h = \frac{1}{a}\Sigma(y\ln y) + \frac{1}{a}\Sigma y - \frac{nb}{a} - \frac{ky_c f(y_c)}{aF(y_c)}\left\{1 + \ln y_c + \frac{f'(y_c)y_c\ln y_c}{f(y_c)} - \frac{f(y_c)y_c\ln y_c}{F(y_c)}\right\} \tag{3.72}$$

$$\partial^2 \ln L/\partial b\partial h = \frac{1}{a}\Sigma(\ln y) + \frac{ky_c(\ln y_c)f(y_c)}{hF(y_c)}\left\{\ln y_c - \psi(b) - \frac{\partial[\ln F(y_c)]}{\partial b}\right\} \tag{3.73}$$

where the terms are as previously defined. The expectation of the (y) term for the partially truncated distribution is as given by equation (3.24). The expected value of the (lny) term has not been derived. Approximations for the (lny) terms may be obtained from the sample. Negative expectations can then be computed through the use of equations (3.68) through (3.73) for completion of the inverse dispersion matrix. Asymptotic variances and covariances of the parameters of the GG distribution for the Type-I censored sample form the elements of the inverted inverse-dispersion matrix. The asymptotic variance of a quantile can be obtained by substitution of the inverted matrix elements and equations (3.65) to (3.67) into equation (3.3e).

In this chapter the asymptotic standard errors of estimate of the parameters and quantiles were developed. Much of these developments represent original contributions.

The derived asymptotic expressions for the LP3 distributions will next be studied using Monte Carlo simulation studies.

CHAPTER 4

SIMULATION STUDIES

Monte Carlo generation provides the only known method for the numerical assessment of the derived asymptotic expressions. Synthetically generated data will be used to determine:

a) the accuracy of the censored model's asymptotic results;

b) the applicability of the Type II censoring model to estimate the asymptotic error of Type I information; and

c) the influence of measurement uncertainty of historic information on the asymptotic error for the Type II model.

Due to the importance of the simulation studies, the Monte Carlo data generation method will be examined.

4.1 The Monte Carlo Generator

Monte Carlo simulation methods consist of the generation of a very large sample of data having assumed statistical properties. These data are considered to be a population, from which samples with a desired number of repetitions can be drawn. This permits an evaluation of proposed procedures.

Several methods of generating gamma type variates are suitable for use in Monte Carlo experiments. Des Groseilliers (1985) provides several algorithms for the generation of gamma variates, defined herein with parameter b. She reviews the work of Tadikamalla and Johnson (1981) who recommend various approaches for generating several gamma variates for a particular parameter value. One of the recommended methods is that of Cheng and Feast (1980) and is applicable for values of b larger than one-half.

Once suitable gamma variates have been obtained, LP3 variates are derived by the change of variate technique of equation (2.62). That is, if y is a gamma variate parameter b and x is a log Pearson Type III variate with parameters a, b, and m, then

$$x = \exp(ay + m) \qquad (2.62a)$$

Another method of generating LP3 variates has appeared in hydrologic literature (Nozdryn-Plotnicki and Watt 1979; Pilon et al. 1987). This approach is a two-step exercise. First, pseudo-random standard normal deviates are obtained. These can be found, for example, from:

a)

$$t = \left[\sum_{i=1}^{v} u_i - (v/2)\right] / \left[\sqrt{v/12}\right] \qquad (4.1)$$

where the u_i's are pseudo-random numbers generated from a uniform distribution over the interval (0,1) and v is the number of pseudo-random uniform numbers generated in order to obtain t, the standard normal deviate. Nozdryn-Plotnicki and Watt (1979) assume a "v" of 12, while Pilon et al. (1987) use a "v" of 100.

b) using an inverse function method based on the work of Hastings (1955). A uniform density on (0,1) can be converted to a standard normal deviate by (Abramowitz and Stegun 1972):

$$t = \frac{W - C_0 + C_1W + C_2W^2}{1 + d_1W + d_2W^2 + d_3W^3} + \varepsilon(u) \tag{4.2}$$

where

$$W = [\ln(1/u^2)]^{\frac{1}{2}} \qquad ; 0 < u \leq .5$$

$$|\varepsilon(u)| < 4.5 \times 10^{-4}$$

and

$$C_0 = 2.515517 \qquad d_1 = 1.432788$$

$$C_1 = .802283 \qquad d_2 = .189269$$

$$C_2 = .010328 \qquad d_3 = .001308$$

The second step in LP3 variate generation is to place the resultant standard normal deviate into the Wilson-Hilferty transformation of equation (3.42a) for an assumed parameter combination.

However, the random number generators are usually of the multiplicative congruential type and have the form

$$u_{i+1} = (c\ u_i + d) \text{ MOD } e \tag{4.3}$$

where c and d are real numbers and are treated as constants, e is a modulus, and u_i is a seed number or initial random number.

The RAN subroutine on a Digital Equipment Corporation (DEC) micro-VAX is used in this research to obtain pseudo-random numbers. RAN's seed is updated automatically and uses the following algorithm to update the seed (DEC 1986):

$$\text{SEED} = 69069\ (\text{SEED}) \dotplus 1(\text{MOD}2^{32}) \tag{4.4}$$

where the SEED is a 32-bit number whose high-order 24 bits are converted to floating point and returned as the result. This implies that random numbers have seven significant figures. As the VAX is a 32-bit computer, the generated sequence of random numbers recycles at $(2^{31}-1)$ or approximately every 2.147 (10^9) values.

When performing Monte Carlo style studies, an unbiased generator which uses the least number of pseudo-random uniform numbers is usually preferred. Such generators can provide a larger number of samples upon which statistical testing of hypotheses can be based.

4.1.1 Testing the Random Generator

The random generator is tested to assure that it is not biased, prior to proceeding with tests regarding the gamma generators. The null hypothesis that the random uniform generator is unbiased is consequently brought forward. The experimental procedure consists of generating 5 million uniform numbers, counting the quantity which exceed prescribed quantiles of .9, .98, .99, .998, and .999. Tests of significance regarding the "biasedness" of the generator can then be performed. Note that these quantiles were selected as flood frequency analysis is particularly interested with the right tail of distributions.

The number of uniform variates from the 5 million sample which exceed the prescribed quantiles are 499.958, 100,381, 50,153, 9,975 and 5,006, respectively. A significance test based on the binomial distribution is used to determine whether or not the generator is biased. The number of variates generated above a certain quantile should follow the binomial distribution.

The test consists of letting P be the probability of a variate greater than or equal to the quantile of interest and n be the sample size, $n=5(10^6)$. The binomial distribution can be approximated in this case by the normal such that the mean number of values above the quantile would be nP and would have a standard deviation of $\sqrt{nPQ}$ where Q is the quantile. Thus, an example is the .9 quantile, where nP is 500,000 with a standard deviation of 671. Hence, the probability that the observed number will deviate from 500,000 by more than 1.96 (671) or 1,315 in either direction is approximately 5%. The generator gives 499,958, thus it is not biased at or above the .9 quantile, for the 5% level.

The aforementioned significance test was conducted as well on the .98, .99, .998, and .999 quantiles resulting with similar conclusions. That is, the random uniform generator is not biased.

4.1.2 Testing the Algorithm of Equation (4.1)

Given that the uniform generator is unbiased, a similar experiment can now be conducted to investigate the ability of equation (4.1) to produce unbiased standard normal deviates. Table 4.1 lists the number of standard normal deviates which were found to exceed selected quantiles. The sample size for this experiment was also set at $5(10^6)$.

The results listed in Table 4.1 indicate that the generator of equation (4.1) with v=12 produces biased standard normal deviates. When the number of uniform variates is increased to 100, the results are generally unbiased. The X^2 goodness-of-fit test can also be applied to the values of Table 4.1. The "v=12" results gives a X^2 of 729, while the "v=100" results yield a X^2 of 6.04. Critical values of the X^2 test are 11.07 at the 5% and 15.09 at the 1% level of significance. X^2 results support the findings of the binomial test.

In summary, the use of only 12 uniform numbers is shown to produce biased standard normal deviates (Table 4.1). However, the bias is almost eliminated when the number of uniform numbers is increased to 100. It should be therefore noted that when the generated data are biased, then the findings from the simulation study are questionable.

Table 4.1 Generation of Standard Normal Deviates using Equation (4.1)

Quantile		Expected Number	Observed Number	95% Lower/Upper
.9	(V=12)	500,000	507,460*	498,686/501,314
	(V=100)		500,371	
.98	(V=12)	100,000	97,639*	99,386/100,614
	(V=100)		99,682	
.99	(V=12)	50,000	46,845*	49,564/50,436
	(V=100)		49,832	
.998	(V=12)	10,000	7,869*	9,804/10,196
	(V=100)		9,792*	
.999	(V=12)	5,000	3,614*	4,861/5,139
	(V=100)		4,906	

*: appears biased at 5%.

4.1.3 Testing of the GT Algorithm

The generation of each standard normal deviate via equation (4.1) requires a large number of uniform variates in order to yield statistically unbiased results. Thus, the generation of gamma variates by approaches which require a lower number of uniform variates to be generated would be preferred. Tadikamalla and Johnson (1981) review fifteen such algorithms which can produce gamma variates. Their results indicate that the "GT" algorithm of Cheng and Feast (1980) is appropriate for the generation of a large number of gamma variates for a particular value of b. Hence, the GT algorithm is one of the preferred approaches for the type of study to be performed in this research.

The GT algorithm was tested for parameter, b, values of 4 and 25 using 5 million replications. Table 4.2 lists the number of values equalling or exceeding selected quantiles. Results indicate that the GT algorithm is not biased. In addition, the GT algorithm required less than $14(10^6)$ uniform random variates in order to produce $5(10^6)$ gamma variates. Based on these results the GT algorithm was selected as the method for generating gamma variates for this research.

Table 4.2 Generation of Gamma Variates using the GT Algorithm of Cheng and Feast (1980)

Quantile (exceedance probability)	Expected Number	Observed Number for parameter 4	25	95% Lower/Upper
.999	5,000	5,038	5,020	4,861/5,139
.99	50,000	50,035	49,969	49,564/50,436
.9	500,000	499,832	500,940	498,686/501,314
.1	500,000	500,563	499,475	498,686/501,314
.02	100,000	99,947	100,090	99,386/100,614
.01	50,000	49,871	50,220	49,564/50,436
.002	10,000	9,934	10,220*	9,904/10,096
.001	5,000	4,917	5,048	4,861/5,139

*: appears biased at 5% level of significance

4.2 Selection of Parameters

Table 4.3 lists the LP3 parameters assumed in the generation of synthetic numbers. The statistics in real and log space are given for the two parameter combinations. It is felt that the assumed LP3 parameters represent typical values in Canada. The first parameter combination is positively skewed in real space but is negatively skewed in log space. This implies that the parent LP3 distribution is negatively skewed and is upper bounded.

Table 4.3 LP3 Parameters and Statistics for Two Samples

Data Set	LP3 Parameters			Natural log Space			Real Space		
	a	b	m	$\bar{X}$	S	CS	$\bar{X}$	S	CS
1	-.06	25	7	5.500	.300	-.4023	255.55	73.829	.4438
2	.05	25	5	6.250	.250	.4000	535.04	143.37	1.3326

$\bar{X}$ = mean; S = standard deviation; CS = coefficient of skew

The second population is positively skewed in both real and log space. The parent LP3 is thus bounded below at m and is unbounded above. The coefficient of variation of the second population in log space is .040; while it is .0545 for the first data set.

Various natural basins in Canada were reviewed in this study to obtain typical flood statistics. These two parameter combinations allow the testing of both the negative and positive cases of the LP3 distribution function.

4.3 Type II Censoring - Results of the Worst Flood in Memory Case

For both data sets of Table 4.3, 500 historic series were extracted from the simulated population. Each series comprised 50 observations such that the largest was the highest observed flow in 100. From the notation of section 3.2.2, the censoring threshold, x_c, was set as the largest in 100, thus n_a was set as one. The number of observed floods less than x_c, which is n_b, was set at 49. And, the number of censored floods, nc, was set at 50. The historic time span was set at 100 and equals the total of n_a, n_b, and n_c.

The values of the .9, .98, .99, and .998 quantiles were computed for each of the two data sets. They were found to be 354, 424, 450, and 503, respectively for the first set and 720, 912, 996, and 1,202, respectively for the second set. These quantiles correspond with the 10-, 50-, 100-, and 500-year floods.

For each historic sample, estimates were made of the LP3 parameters a, b, and m. Table 4.4 summarizes the results of the analysis of the 500 samples. Table 4.4 shows that the estimated flood quantiles are slightly biased. The significance of the bias can be determined following

Table 4.4 Summary of Parameters and Quantiles Estimated from 500 Sequences using Type II Censoring

Parameter or Event	Population Value	Mean Estimate	Bias %	Standard Deviation	RMS[1]
(a) Data Set 1					
a	-.06	-.0655	9.2	.0529	-
b	25	77.214	209	168	-
m	7	6.722	4.0	2.21	-
Q_{10}	354	354	.07	15.2	15.17
Q_{50}	424	423	.23	25.3	25.34
Q_{100}	450	449	.28	32.3	32.28
Q_{500}	503	502	.09	53.3	53.26
(b) Data Set 2					
a	.05	.055	10.4	.0404	-
b	25	95.8	283	229.	-
m	5	5.01	.26	2.08	-
Q_{10}	720	720	.04	37.0	36.92
Q_{50}	912	917	.57	75.3	75.38
Q_{100}	996	1,006	.95	99.2	99.53
Q_{500}	1,202	1,229	2.28	174.	176.2

1: RMS implies root mean square error

the procedure adopted by Lowery and Nash (1970). Basically, Student's t-test is applied to the means, such that:

$$t = \sqrt{N}\,(Z-\bar{Z})/\hat{\sigma}_z \qquad 4.5$$

where $\bar{Z}$ is the mean of N estimates of the population statistics $\bar{Z}$, $\hat{\sigma}_z$ is the estimated standard deviation of the N estimates about Z, and t is distributed like Student's t with (N-2) degrees of freedom.

Table 4.5 summarizes the application of the t-test on the estimate of the quantiles for the two data sets. At the 5% level of significance and with the appropriate degrees of freedom, the tabulated t is 1.96, while the tabulated t is 2.576 at the 1% level. At this latter level, only the 500-year flood estimate appears statistically biased. Table 4.4 lists the bias in percent for the quantiles. The bias of the 500-year flood for the second data set is less than 2.5%, while the remaining quantiles are less than 1% biased. In hydrometric terms, a 2.5% difference in means may not be practically different.

The asymptotic standard error of estimate of the various quantiles can be estimated using the expressions presented in Section 3.5.2. Appendix A lists and documents a computer program which can be used to estimate such errors in the historic and the conventional cases. In order to assess the asymptotic standard error of estimate for the worst-flood in memory case, a standard normal deviate corresponding with its exceedance probability must be assigned to the flood. Rank-order statistics are used to obtain the exceedance probability. The Cunnane (1978) formulae is used

Table 4.5 Values of Student's t for 500 Type II Samples

	Design Flood Q_{10}	Q_{50}	Q_{100}	Q_{500}
Data Set 1	-0.37	0.87	0.86	0.19
Data Set 2	0.19	-1.54	-2.13[1]	-3.48[1]

1: calculated t exceeds tabulated t at the 5% level of significance for a two-tailed test

Table 4.6 Root Mean Square Errors for Type II Censoring

		Design Flood Q_{10}	Q_{50}	Q_{100}	Q_{500}
Data Set 1	Sample	15.173	25.344	32.278	53.255
	asymptotic	15.682	24.879	31.928	53.275
Data Set 2	Sample	36.919	75.379	99.534	176.220
	asymptotic	37.080	77.264	102.376	179.484

Table 4.7 F Ratios of Asymptotic Error and Mean Square Error for Type II Censoring

	Design Flood Q_{10}	Q_{50}	Q_{100}	Q_{500}
Data Set 1	1.068	1.038	1.022	1.001
Data Set 2	1.009	1.051	1.058	1.037

as an approximation in obtaining the return period, T, of the worst flood in memory as

$$T = \frac{YT + .2}{m - .4}$$

where YT is the historic time span and m is the rank of the flood in question which in this case is one. Thus, the return period of the worst flood in one hundred years is estimated to be 167 years. Its exceedance probability is the inverse of the return period and is 0.00599. This corresponds with a standard normal deviate of 2.51.

Table 4.6 lists the computed asymtotic standard errors of estimate for the selected design floods. It also contains the computed root mean square errors from the 500 samples. The root mean square error represents the standard deviation of the computed design floods about their population's mean.

Lowery and Nash (1970) and Condie and Lee (1982) apply the F-test to the mean square errors. The null hypothesis is that the asymptotic results are like those obtained from the Monte Carlo experiment. The F ratio is always computed as the quotient of the larger variance divided by the smaller. The degrees of freedom associated with the Monte Carlo experiment are approximately 500, while the degrees of freedom for the asymptotic results are taken as infinity. The critical value of F at the 5% level of significance when $\nu_1=500$ and $\nu_2=\infty$ is 1.15, while when $\nu_1=\infty$ and $\nu_2=500$ the critical F is 1.12. Any computed values of F greater than the appropriate critical value infers that the two measures of variability are significantly different. Values of F greater than the critical values can occur by chance only with a probability of less than 5%.

Table 4.7 lists the computed F statistic for each of the two data sets and the various design floods. The asymptotic results do not appear to be significantly different than the Monte Carlo results. Thus, the theoretical expressions developed herein appear to accurately estimate the asymptotic standard error of estimate of the quantiles for the worst flood in memory case - Type II censoring.

4.3.1 Distribution of the Quantiles

It is commonly assumed that the estimates of the quantiles are normally distributed (Kite 1975; Kendall and Stuart 1979; Ashkar and Bobée 1988). This assumption facilitates the calculation of confidence limits for selected quantiles. The Monte Carlo experiment provides samples of quantile estimates which can be analyzed to determine the adequacy of the normality assumption.

Figures 4.1 through 4.8 show the frequency histograms of the estimated 10-, 50-, 100-, and 500-year design floods for the two data sets. In addition, these figures display the expected number of observations for each class interval had the data been normally distributed with a mean and standard deviation of the quantile estimates. Table 4.8 lists these and other pertinent statistics on the 500 members of each quantile.

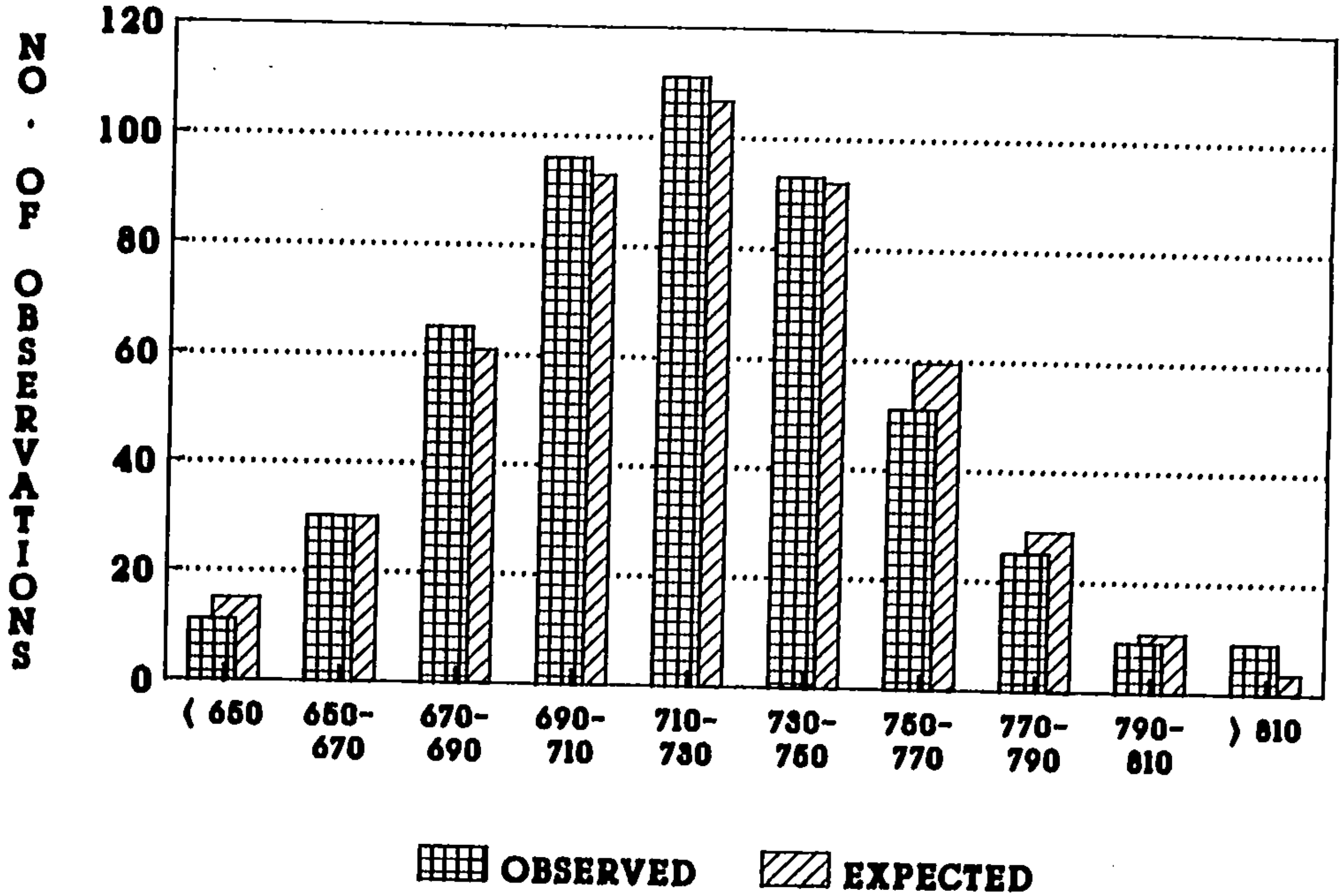


Figure 4.1: Frequency Histograms of the 500 10-Year Design Floods Estimated by Type II Censoring for Data Set 1 (Table 4.3) and the Expected Number if Normally Distributed

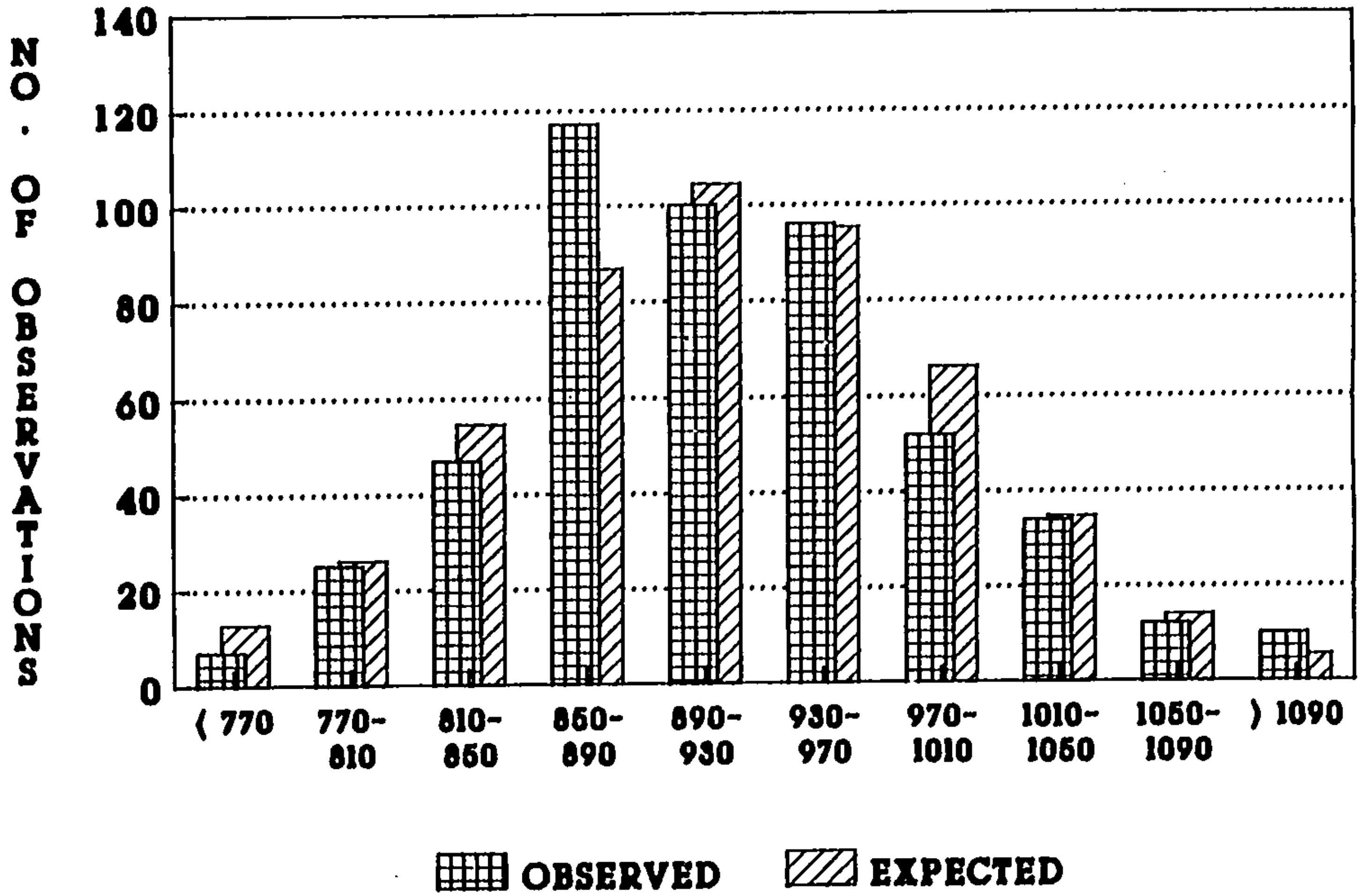


Figure 4.2: Frequency Histograms of the 500 50-Year Design Floods Estimated by Type II Censoring for Data Set 1 (Table 4.3) and the Expected Number if Normally Distributed

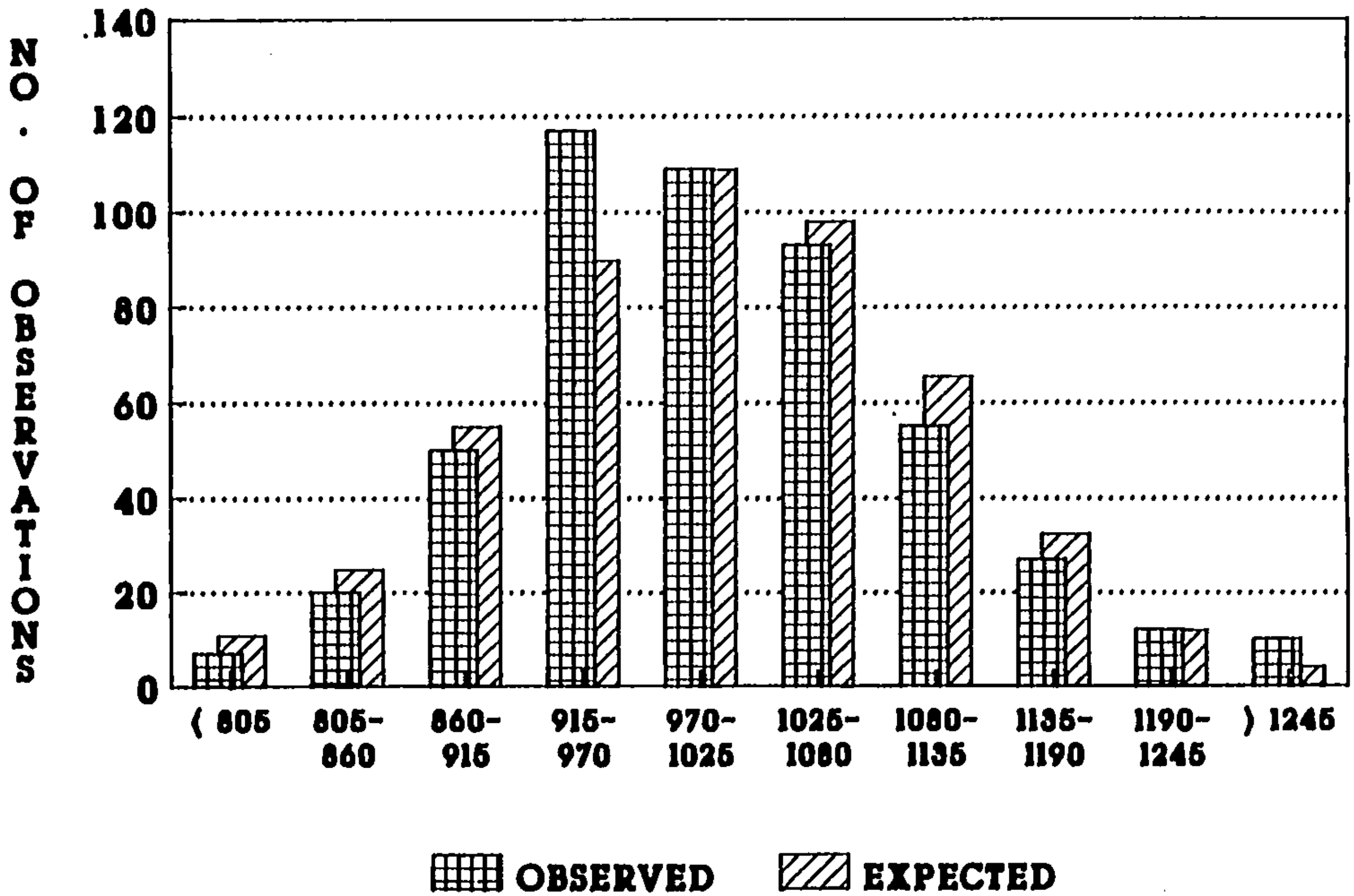


Figure 4.3: Frequency Histograms of the 500 100-Year Design Floods Estimated by Type II Censoring for Data Set 1 (Table 4.3) and the Expected Number if Normally Distributed

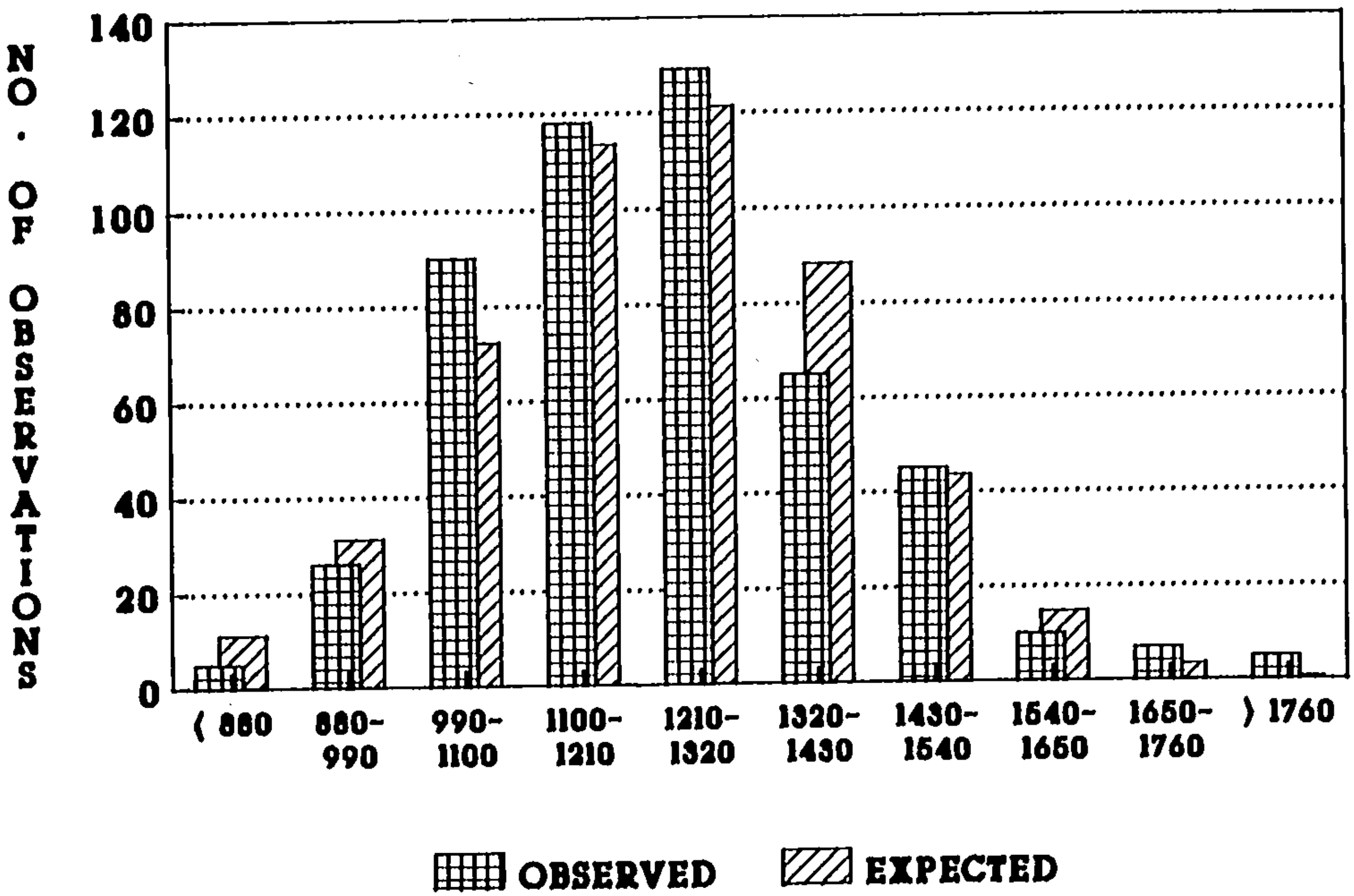


Figure 4.4: Frequency Histograms of the 500 500-Year Design Floods Estimated by Type II Censoring for Data Set 1 (Table 4.3) and the Expected Number if Normally Distributed

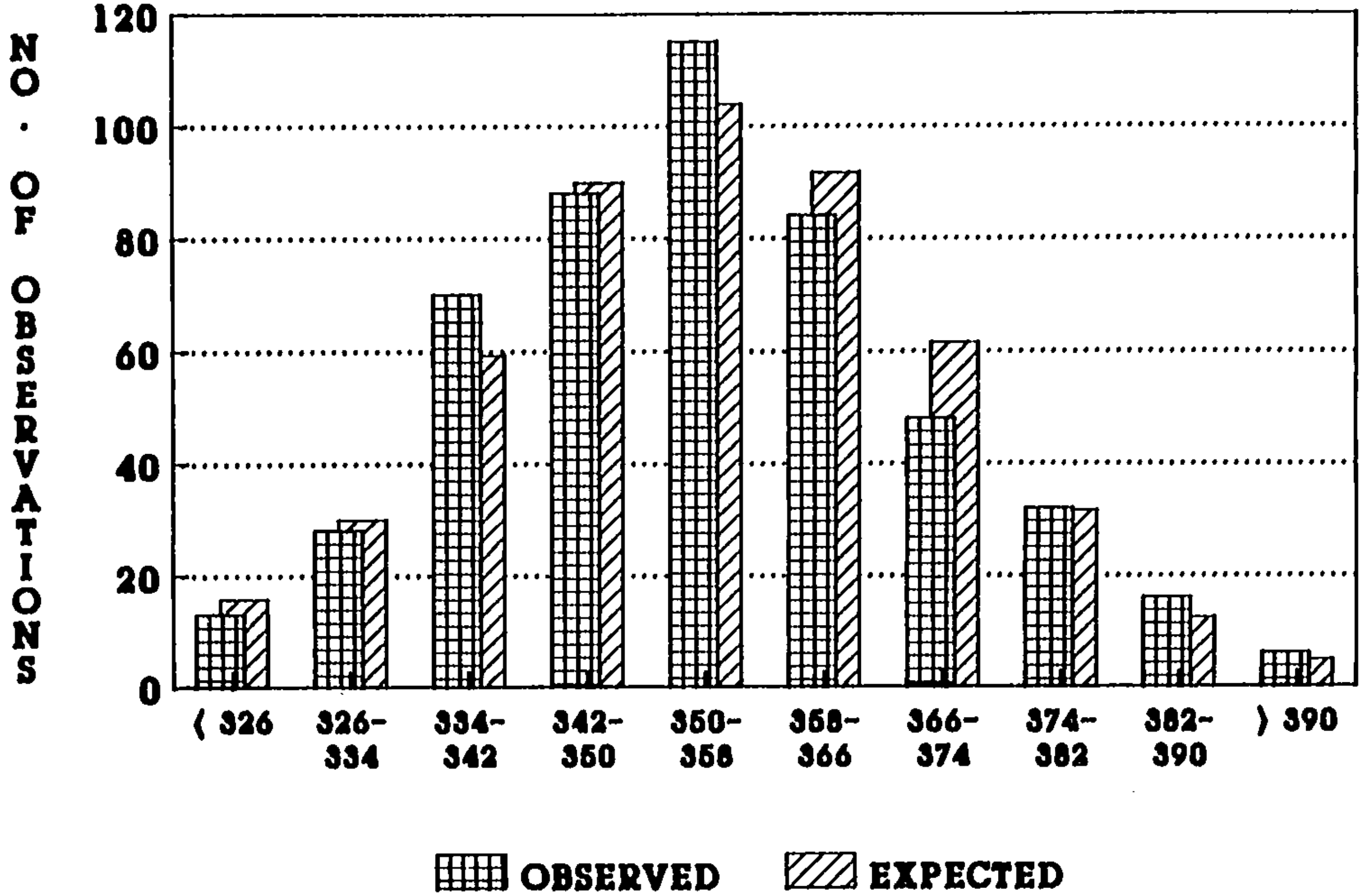


Figure 4.5: Frequency Histograms of the 500 10-Year Design Floods Estimated by Type II Censoring for Data Set 2 (Table 4.3) and the Expected Number if Normally Distributed

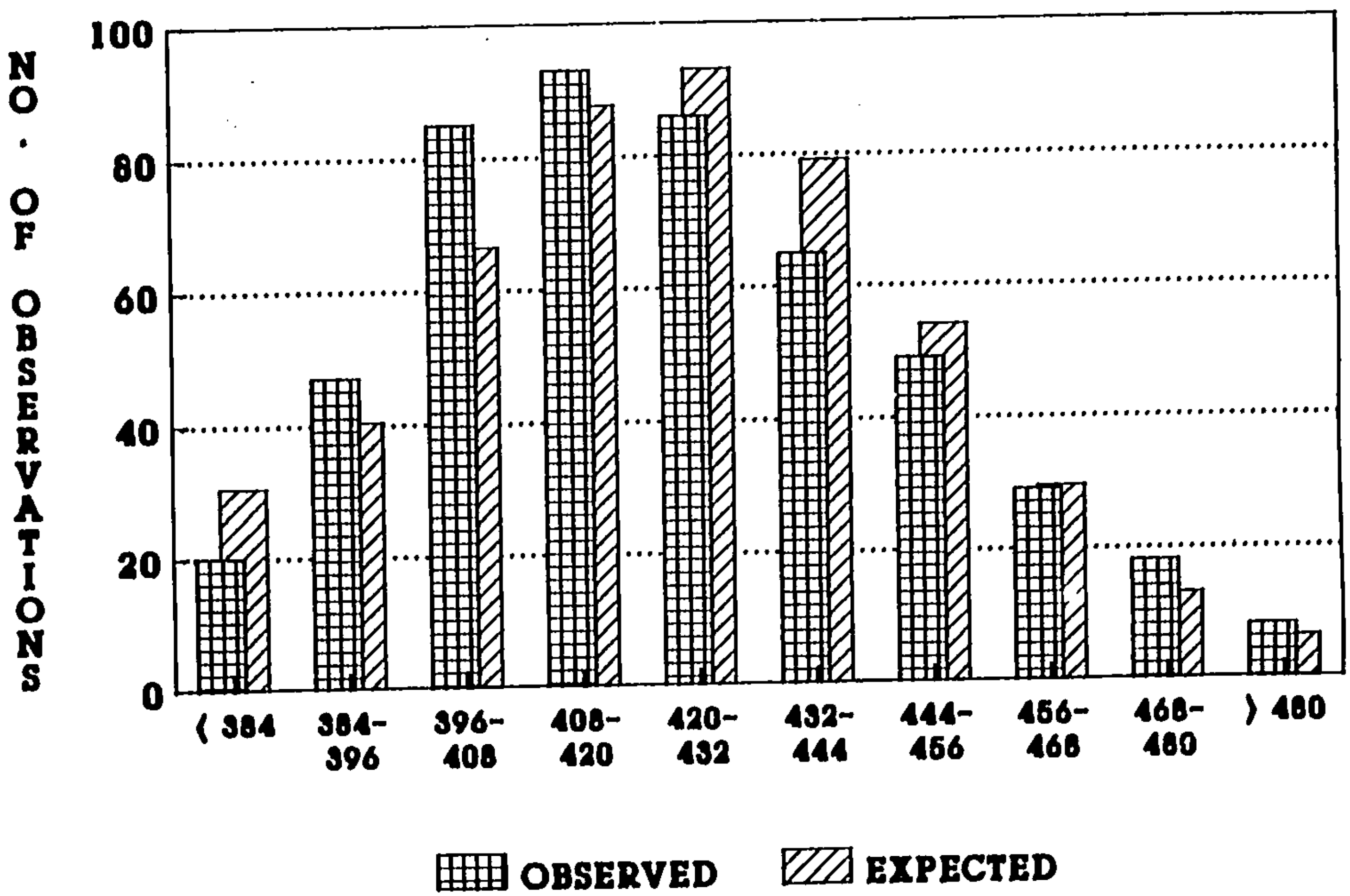


Figure 4.6: Frequency Histograms of the 500 50-Year Design Floods Estimated by Type II Censoring for Data Set 2 (Table 4.3) and the Expected Number if Normally Distributed

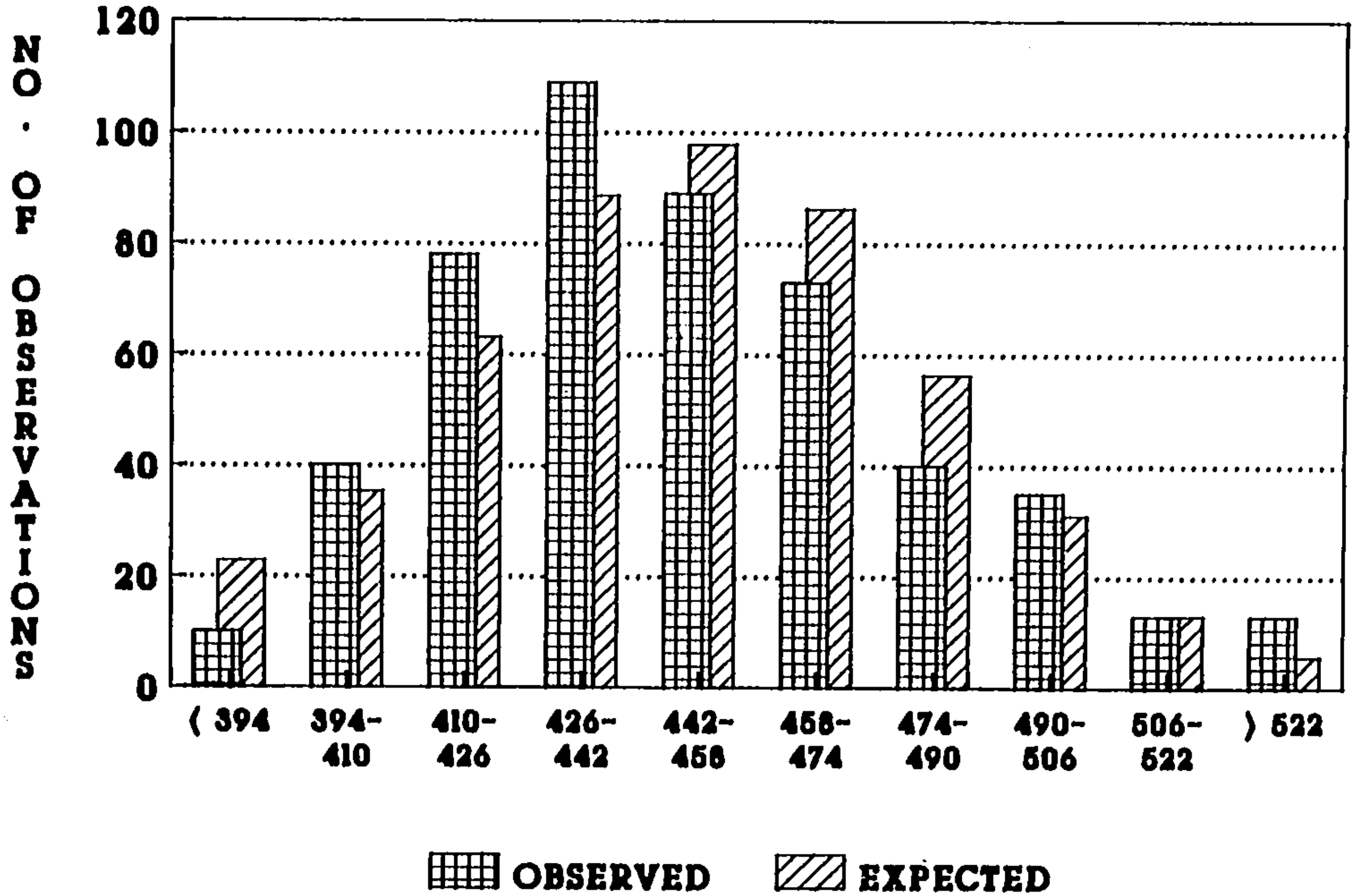


Figure 4.7: Frequency Histograms of the 500 100-Year Design Floods Estimated by Type II Censoring for Data Set 2 (Table 4.3) and the Expected Number if Normally Distributed

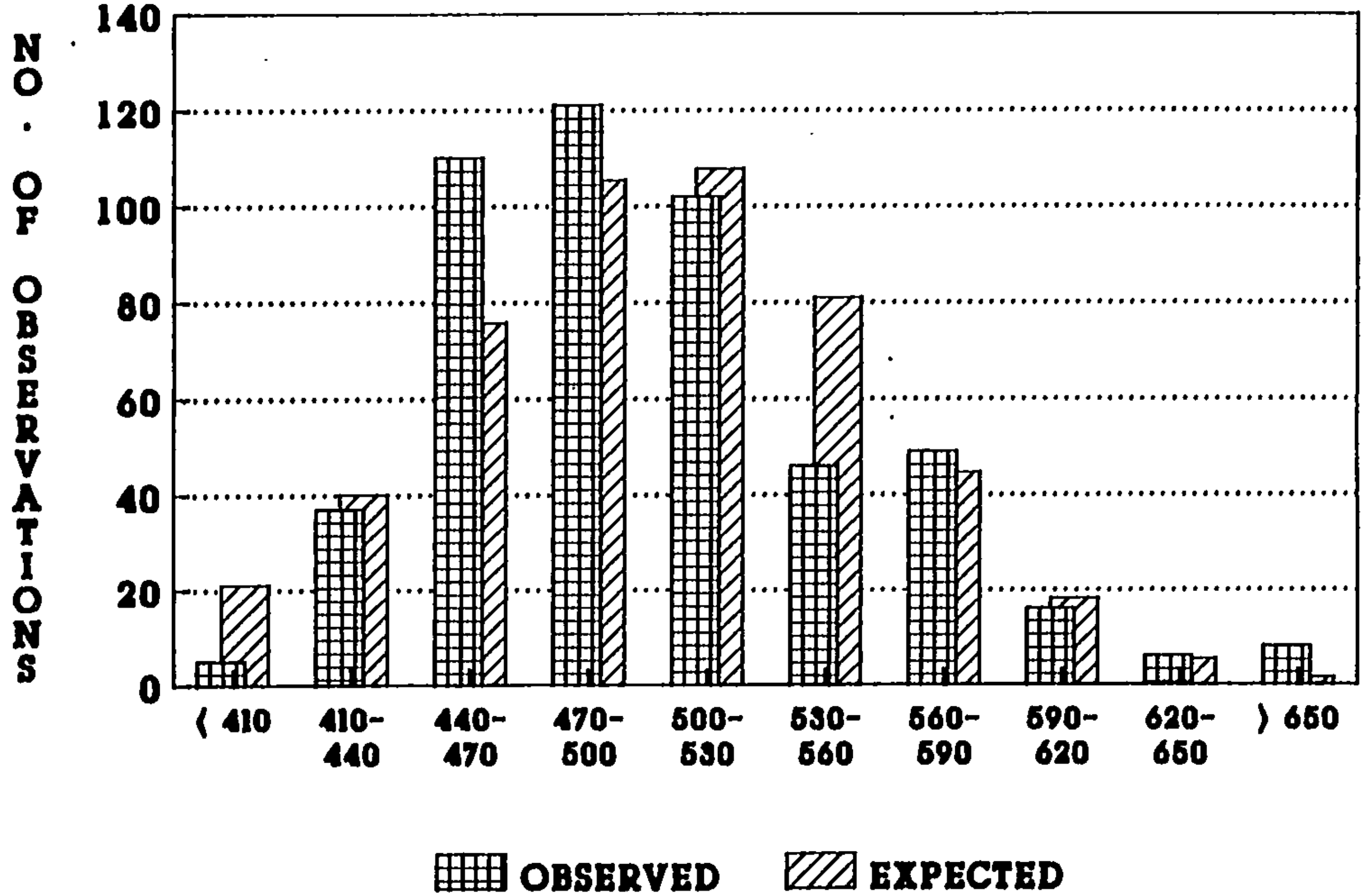


Figure 4.8: Frequency Histograms of the 500 500-Year Design Floods Estimated by Type II Censoring for Data Set 2 (Table 4.3) and the Expected Number if Normally Distributed

Table 4.8 Statistics of the 500 Estimated Design Floods for the Type II Case

Statistic	Design Flood Q_{10}	Q_{50}	Q_{100}	Q_{500}
Data Set 1				
Mean	354.3	423.2	448.6	502.2
Standard Deviation	15.2	25.3	32.3	53.3
Coefficient of Skew	.331	.453	.555	.877
Kurtosis	2.98	3.19	3.26	4.03
Coefficient of Excess	-.02	.19	.26	1.03
Data Set 2				
Mean	719.6	916.9	1,005.6	1,229.0
Standard Deviation	37.0	75.3	99.2	174.2
Coefficient of Skew	0.506	0.571	0.592	0.664
Kurtosis	4.13	3.94	3.84	3.79
Coefficient of excess	1.13	0.94	0.84	0.79

Table 4.9 Chi-squared Test for Goodness of Fit of Quantile Estimates of Type II Censoring to Normal

	Design Flood Q_{10}	Q_{50}	Q_{100}	Q_{500}
Data Set 1	8.80(ν=7)[1]	16.16(ν=7)[2]	33.34(ν=7)[2]	77.68(ν=7)[2]
Data Set 2	4.40(ν=6)	21.67(ν=7)[2]	22.70(ν=6)[2]	16.16(ν=5)[2]

1: ν implies the degrees of freedom for the χ^2 test.
2: computed χ^2 is greater than the tabulated χ^2 at the 5% level of significance

Several statistical tests can be used to verify the hypothesis that the distribution of the quantile estimates is normal. Under the normality assumption, the coefficients of skew and excess are approximately distributed as normal variables as N(0;6/n) and N(0;24/n), respectively (Phien et al. 1982). At the 5% level of significance, the coefficient of skew and excess should not deviate from zero by more than .21 and .43, respectively. A comparison of these values with the values of Table 4.8 indicate that based on sample skewness the null hypothesis that the quantile estimates are normally distributed can be rejected for all quantiles tested and for both Data Set 1 and 2. A comparison of the range of the coefficient of excess with computed values of Table 4.8 shows that the null hypothesis, which is that the quantiles are normally distributed, can be rejected. The coefficient of excess increases for Data Set 1 as the quantile increases, with the null hypothesis being rejected at the 500-year level.

The aforementioned results for the standard moment ratio tests indicate that the estimates of the quantiles for the Type II censoring case do not appear to be normally distributed. To complement these results, the χ^2 goodness-of-fit test is applied to the two data sets. Table 4.9 lists the computed χ^2 statistic for each selected quantile for both data sets. The calculated χ^2 statistic exceeds the tabulated value at the 5% level of significance for the 50-, 100-, and 500-year design floods. Only the distribution of the 10-year design flood could not be distinguished from being normal.

Results of χ^2 goodness-of-fit data confirm the results of the standard moment ratio tests. That is, the quantile estimates do not appear to be normally distributed. This departure from normality increases with the quantile's exceedance probability.

This discovery that the quantiles are not normally distributed is of great significance in the practical applications of flood frequency analysis when computing confidence levels. Further studies are thus urgently warranted to determine the consequences of the violation of the normality assumption.

4.4 Type I Censoring - Results of the Out-of-Bank Flooding Case

Five hundred Type I censored samples were constructed for both data sets described in Table 4.3. Each sample was derived from one hundred observations. The first fifty observations were scanned for values greater than or equal to the censoring threshold. All values of this nature were set aside, with the total number being noted. This first fifty represents the marker data. The second fifty observations were taken as the systematic portion of the record. These fifty were scanned for values greater than or equal to the threshold, with the number being noted.

The censoring threshold, x_c, was set at the 10-year flood for this exercise. As the probability of a flood equalling or exceeding the 10-year flood is 0.1, the expected number of floods in a sample of 100 to equal or exceed the 10-year flood is 10. Five of the ten would be expected in the marker portion, with the remaining five being from the systematic record portion. It then follows that the number of fully defined floods less than the threshold would be 45, as would the number of censored floods. Thus n_b and n_c would be 45. The summation of n_a, n_b, and n_c would again total the historic time span of 100 years.

Table 4.10 lists various statistics for the parameter of the LP3 distribution estimated from the 500 Type I censored samples. Statistics on four preselected flood quantiles are also included. The bias is statistically assessed by use of the t-test as described in Section 4.3. Table 4.11 lists the calculated t values for each flood quantile and data set. The higher quantiles for the first data set display significant bias, while only the smallest quantile of the second data set appears biassed. Table 4.10 lists the percent bias to be less than 1.5. However, due to the large number of replications in the Monte Carlo study, a bias of this magnitude can be shown to be statistically significant. Thus, maximum likelihood estimators of flood quantiles from Type I censored samples tend to be biassed. This bias tends to be less than 1% and as such may be regarded as being practically insignificant.

The asymptotic standard error of estimate of the various quantiles derived for the Type I censored case can, as well, be estimated using the expressions presented in Section 3.5.2. The computer program of Appendix A, mentioned in the previous section, can also be used to estimate the asymptotic standard error for the Type I censored sample. Table 4.12 lists the computed asymptotic standard error of estimate for the selected design floods corresponding to both data sets. Table 4.12 also contains the computed root mean square error from the 500 Type I censored samples.

Table 4.10 Summary of Parameters and Quantiles Estimated from 500 Sequences using Type I Censoring

Parameter or Event	Population Value	Mean Estimate	Bias %	Standard Deviation	RMS[1]
(a) Data Set 1					
a	-.06	-.0714	19	.0494	-
b	25	58.738	135	125	-
m	7	6.960	0.6	1.67	-
Q_{10}	354	353.4	0.2	12.9	12.86
Q_{50}	424	420.5	0.8	21.9	22.19
Q_{100}	450	444.9	1.1	28.2	28.60
Q_{500}	503	495.7	1.5	47.2	47.62
(b) Data Set 2					
a	.05	.0519	3.8	.0397	-
b	25	56.912	108	104	-
m	5	5.300	6.0	1.60	-
Q_{10}	720	717.3	0.4	29.7	29.76
Q_{50}	912	910.4	0.2	62.5	62.47
Q_{100}	996	996.6	0.1	84.0	83.93
Q_{500}	1,202	1,212.4	0.9	151.9	152.13

1: RMS implies root mean square error

Table 4.11 Values of Student's t for 500 Type I Samples

	Design Flood			
	Q_{10}	Q_{50}	Q_{100}	Q_{500}
Data Set 1	1.136	3.734[1]	3.895[1]	3.282[1]
Data Set 2	1.993[1]	0.465	-0.138	-1.585

1: calculated t exceeds tabulated t at the 5% level of significance for a two-tailed test.

Table 4.12 Root Mean Square Errors for Type I Censoring

		Design Flood			
		Q_{10}	Q_{50}	Q_{100}	Q_{500}
Data Set 1	Sample	12.86	22.20	28.60	47.62
	Asymptotic	12.73	21.40	27.46	45.16
Data Set 2	Sample	29.76	62.47	83.93	152.13
	Asymptotic	30.37	60.18	78.42	133.92

Table 4.13 lists the computed F statistic for each quantile and data set as found in Table 4.12. Only the computed F statistic for the .998 quantile of the second data set exceeds the tabulated critical value of F at the 5% level of exceedance. Generally, the asymptotic error tends not to be significantly different than that found from the Monte Carlo studies.

4.4.1 Distribution of the Quantiles

The distribution of the estimates of the quantiles is shown in Figures 4.9 through 4.16. These figures are the histograms of the estimated 10-, 50-, 100-, and 500-year design floods for both data sets. These figures also display the expected number of observations in each class interval had the estimates being normally distributed. Table 4.14 lists the mean and standard deviation of the quantile estimates as well as other pertinent statistics.

As in section 4.3, an assessment of the normality of each quantile's estimates can be made by the moment ratio tests. Given that the sample size is identical with that of Section 4.3, the coefficient of skew and excess should not deviate from zero by more than .21 and .43, respectively. The coefficients of skew as found in Table 4.14 exceed .21 for all but the 10-year flood. Similar results are found with the coefficient of excess for the second data set, while for the first data set only the 500-year flood estimate is significantly different than zero.

Table 4.13 F Ratios of Asymptotic Error and Mean Square Error for Type I Censoring

	Design Flood			
	Q_{10}	Q_{50}	Q_{100}	Q_{500}
Data Set 1	1.021	1.076	1.085	1.112
Data Set 2	1.041	1.078	1.145	1.290*

*: Calculated value of F exceeds the critical value of F at the 5% level of significance

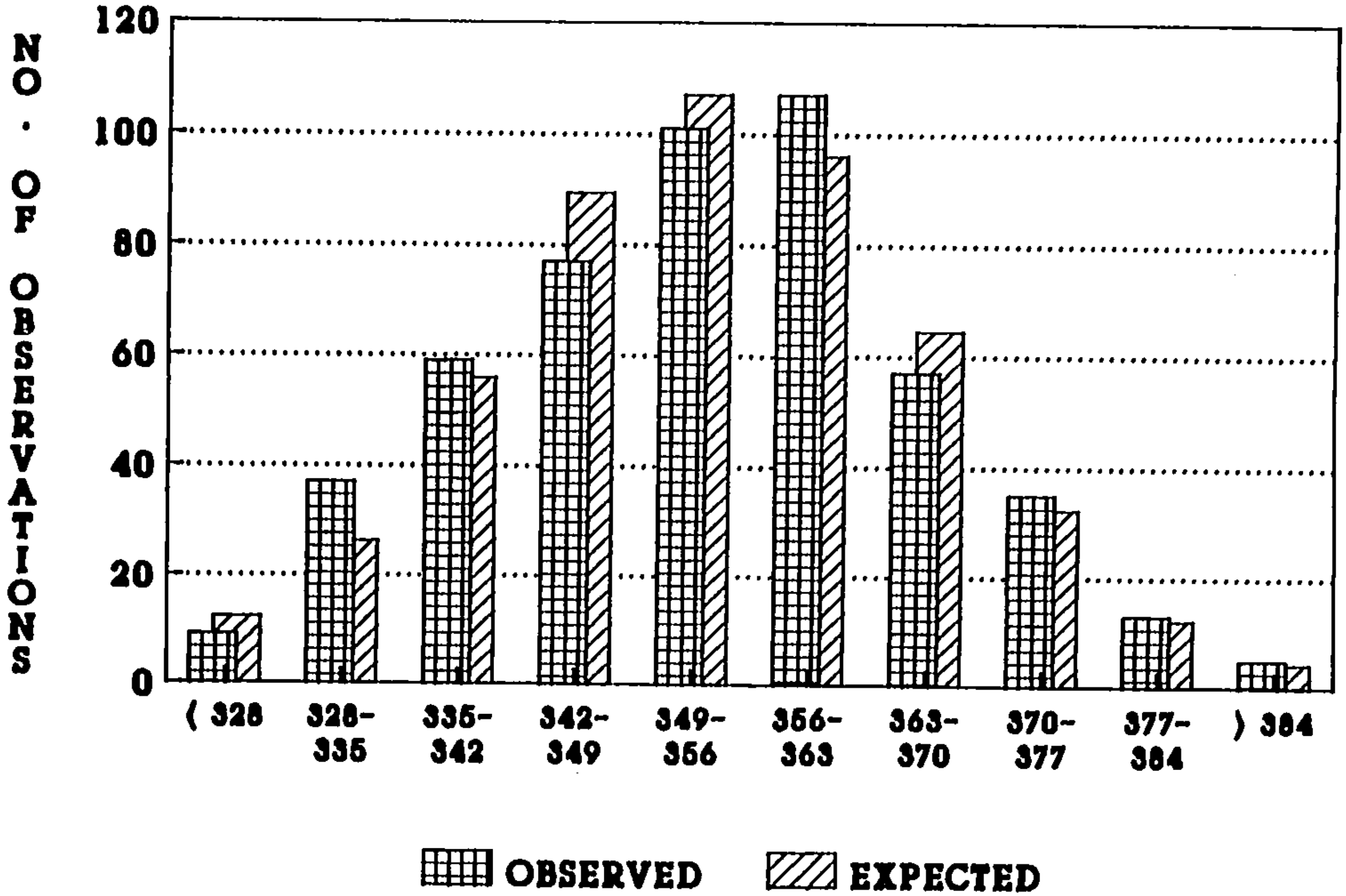


Figure 4.9: Frequency Histograms of the 500 10-Year Design Floods Estimated by Type I Censoring for Data Set 1 (Table 4.3) and the Expected Number if Normally Distributed

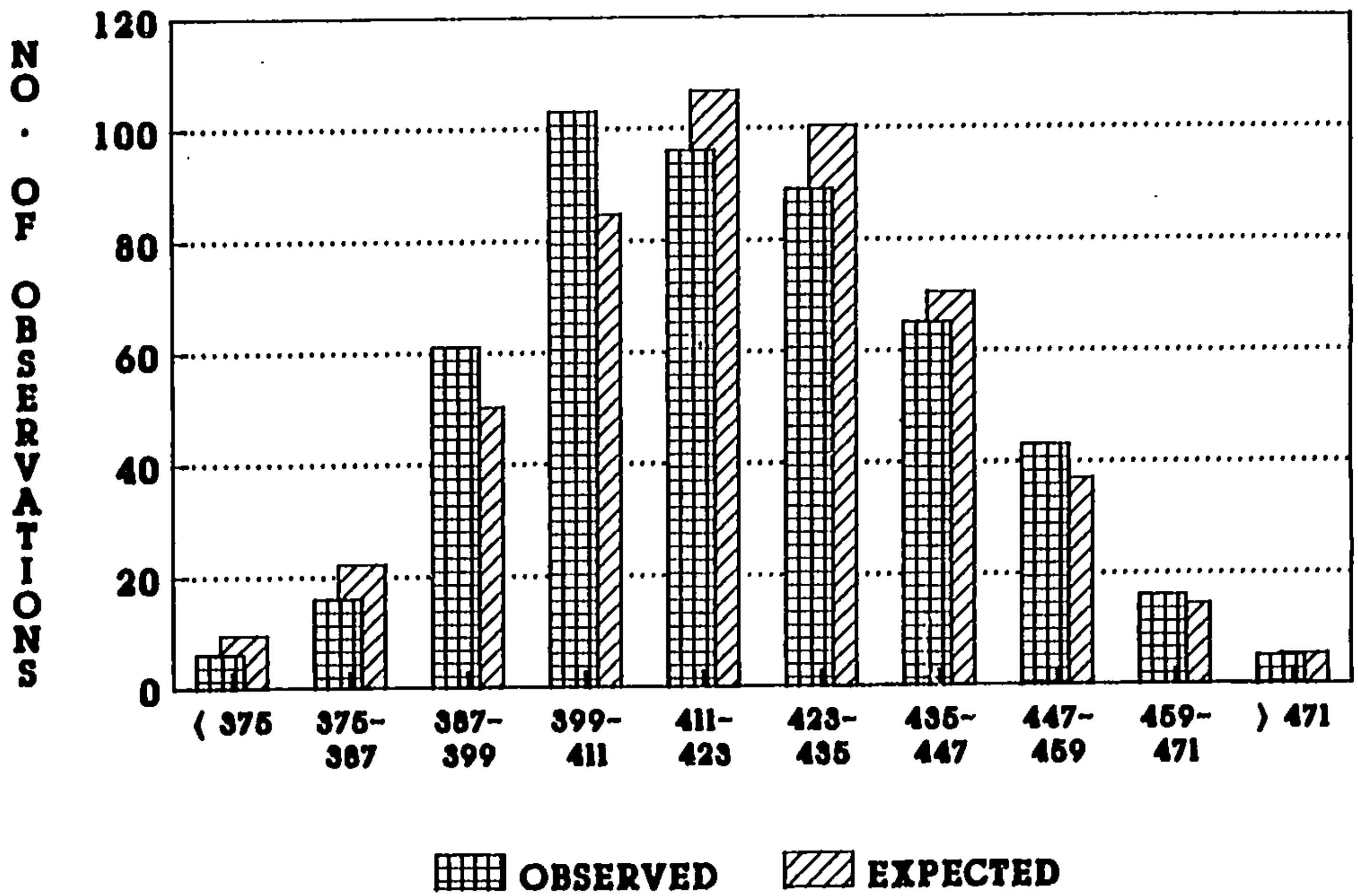


Figure 4.10: Frequency Histograms of the 500 50-Year Design Floods Estimated by Type I Censoring for Data Set 1 (Table 4.3) and the Expected Number if Normally Distributed

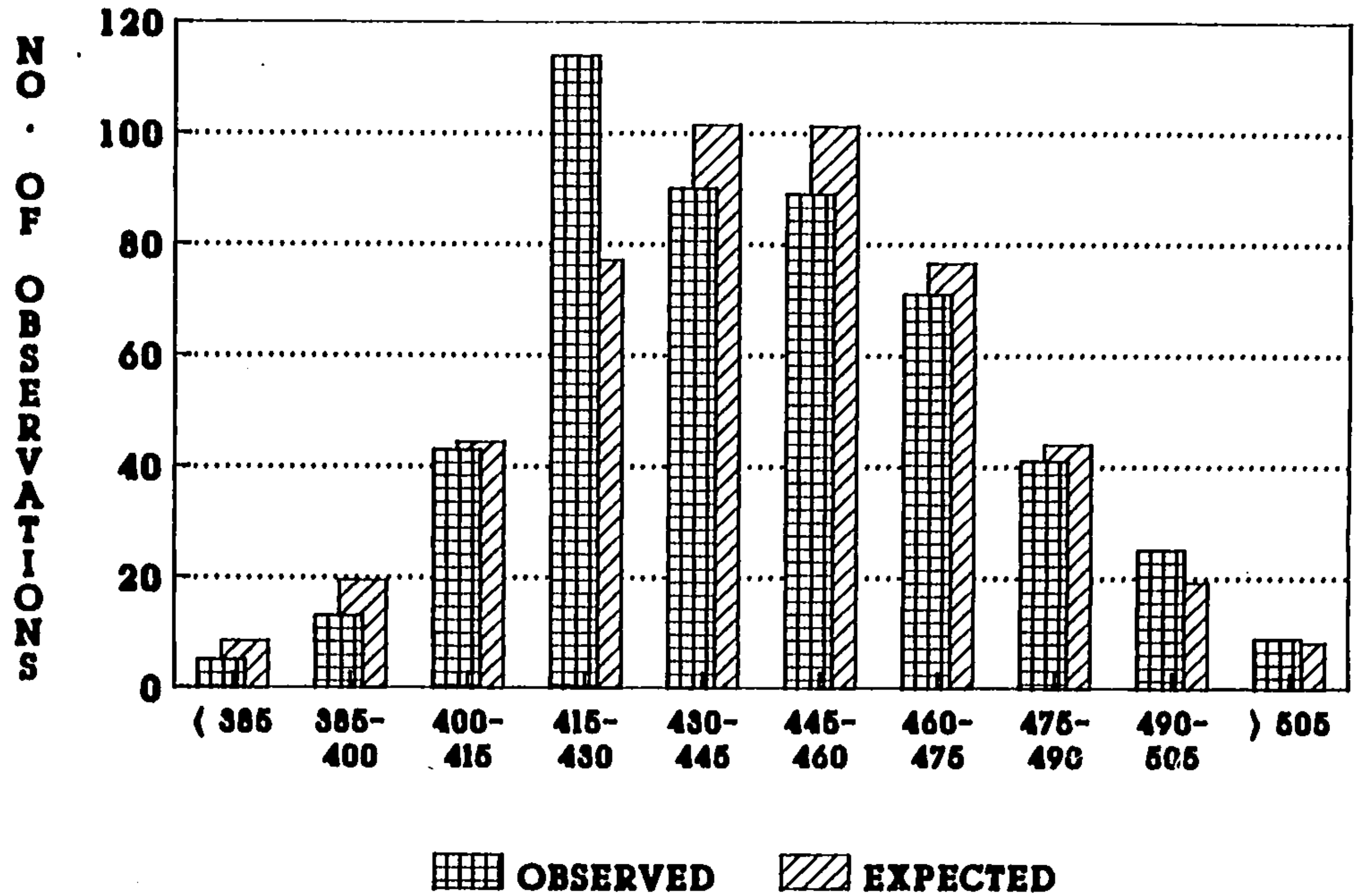


Figure 4.11: Frequency Histograms of the 500 100-Year Design Floods Estimated by Type I Censoring for Data Set 1 (Table 4.3) and the Expected Number if Normally Distributed

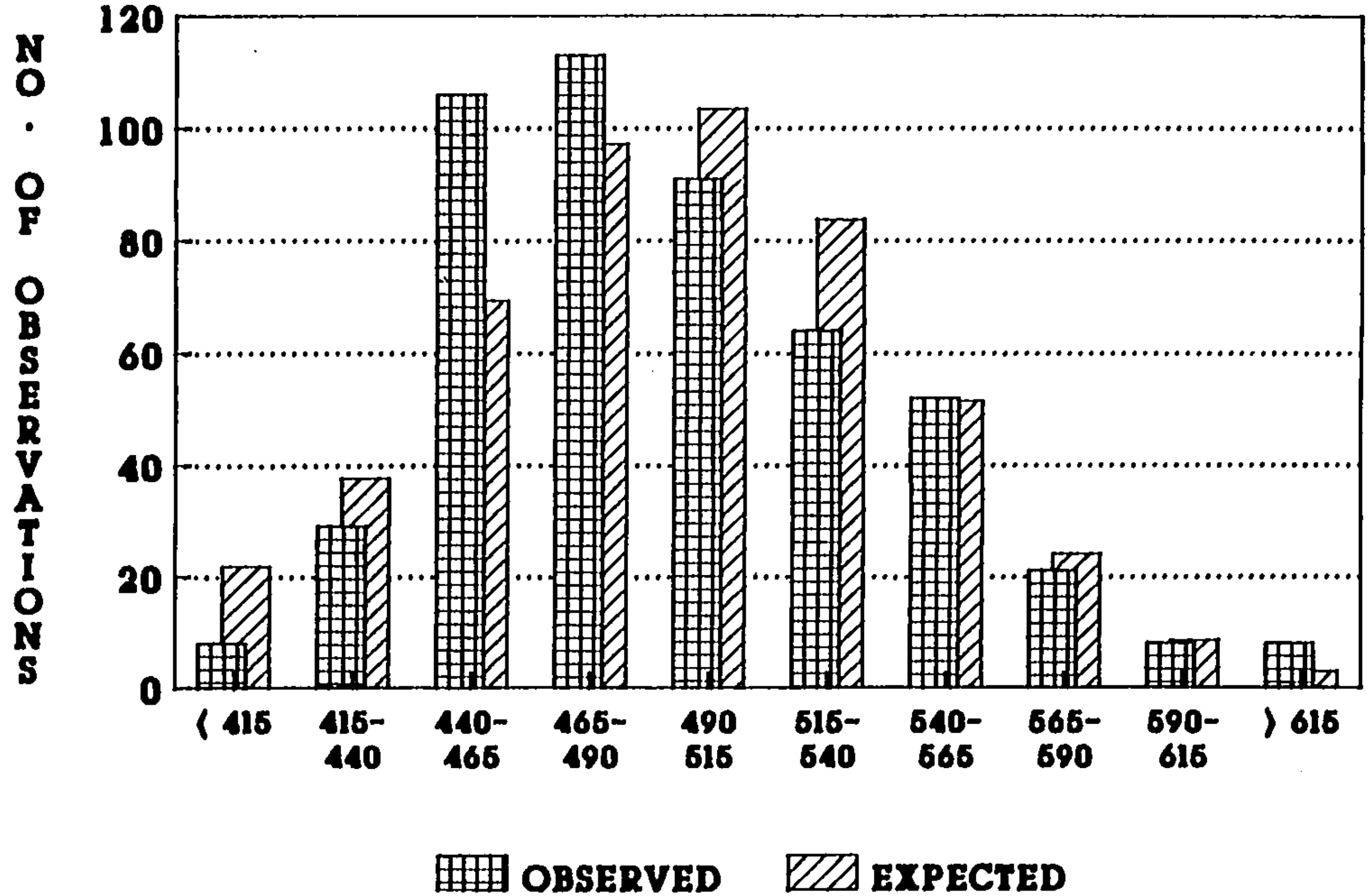


Figure 4.12: Frequency Histograms of the 500 500-Year Design Floods Estimated by Type I Censoring for Data Set 1 (Table 4.3) and the Expected Number if Normally Distributed

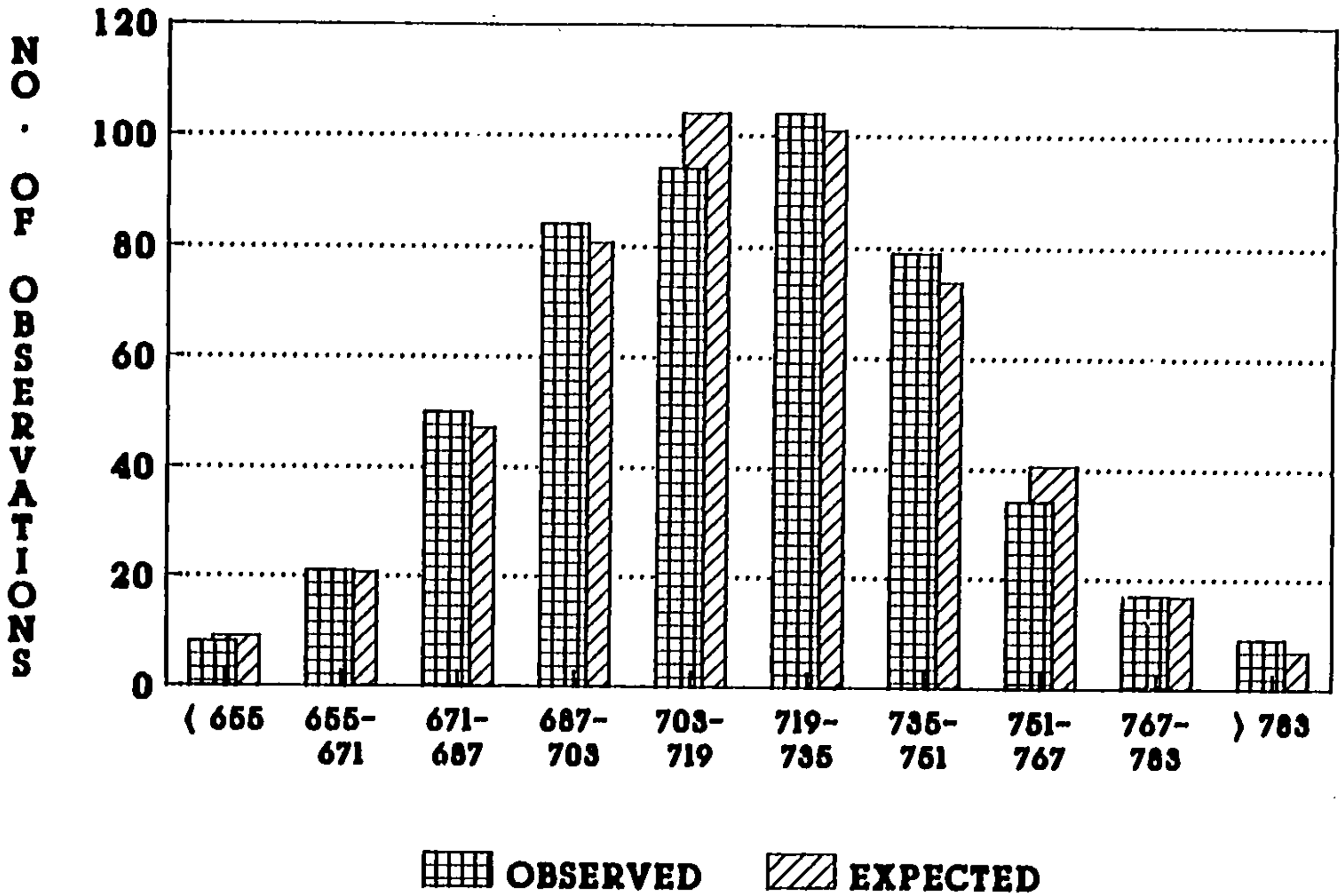


Figure 4.13: Frequency Histograms of the 500 10-Year Design Floods Estimated by Type I Censoring for Data Set 2 (Table 4.3) and the Expected Number if Normally Distributed

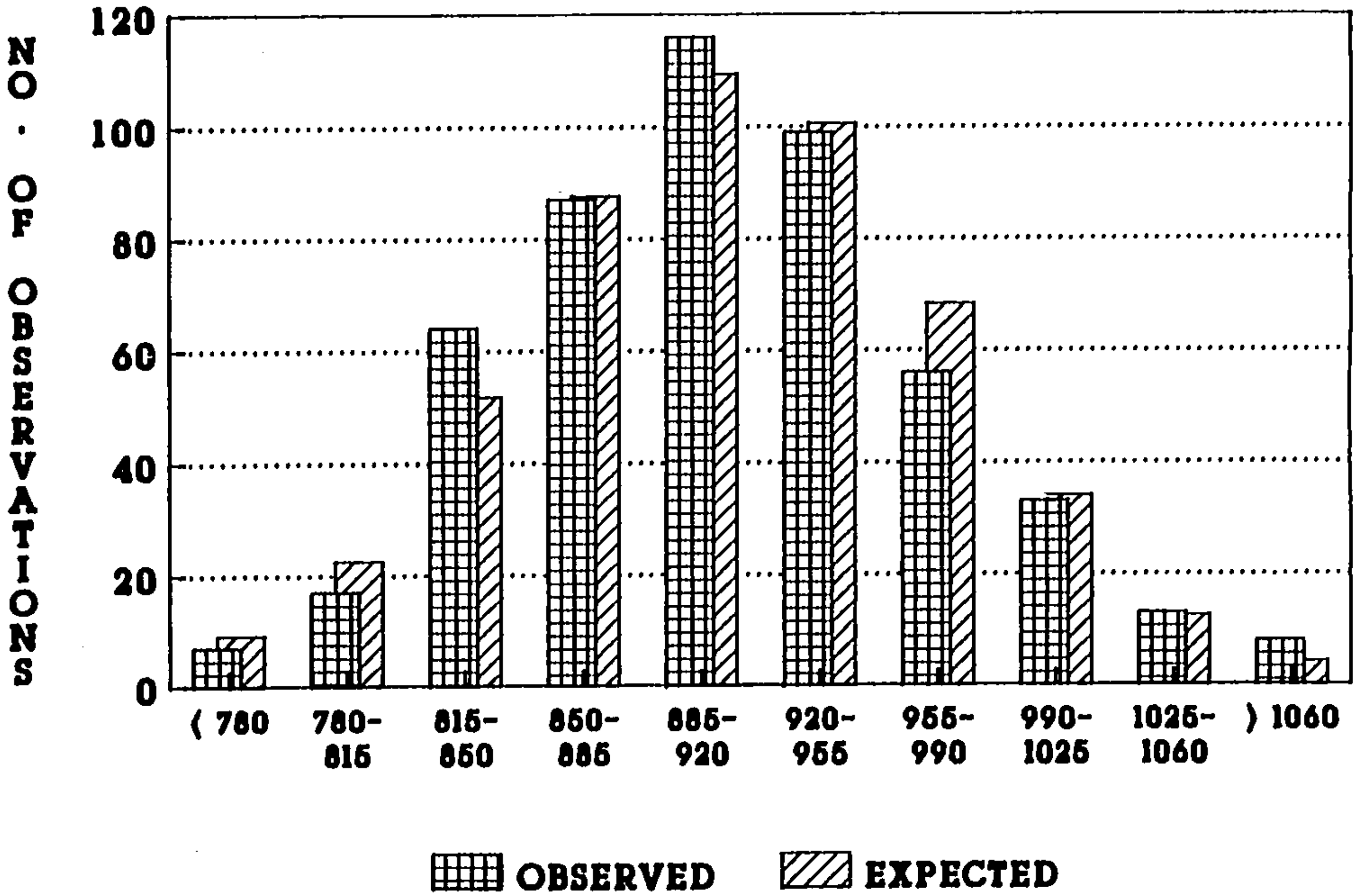


Figure 4.14: Frequency Histograms of the 500 50-Year Design Floods Estimated by Type I Censoring for Data Set 2 (Table 4.3) and the Expected Number if Normally Distributed

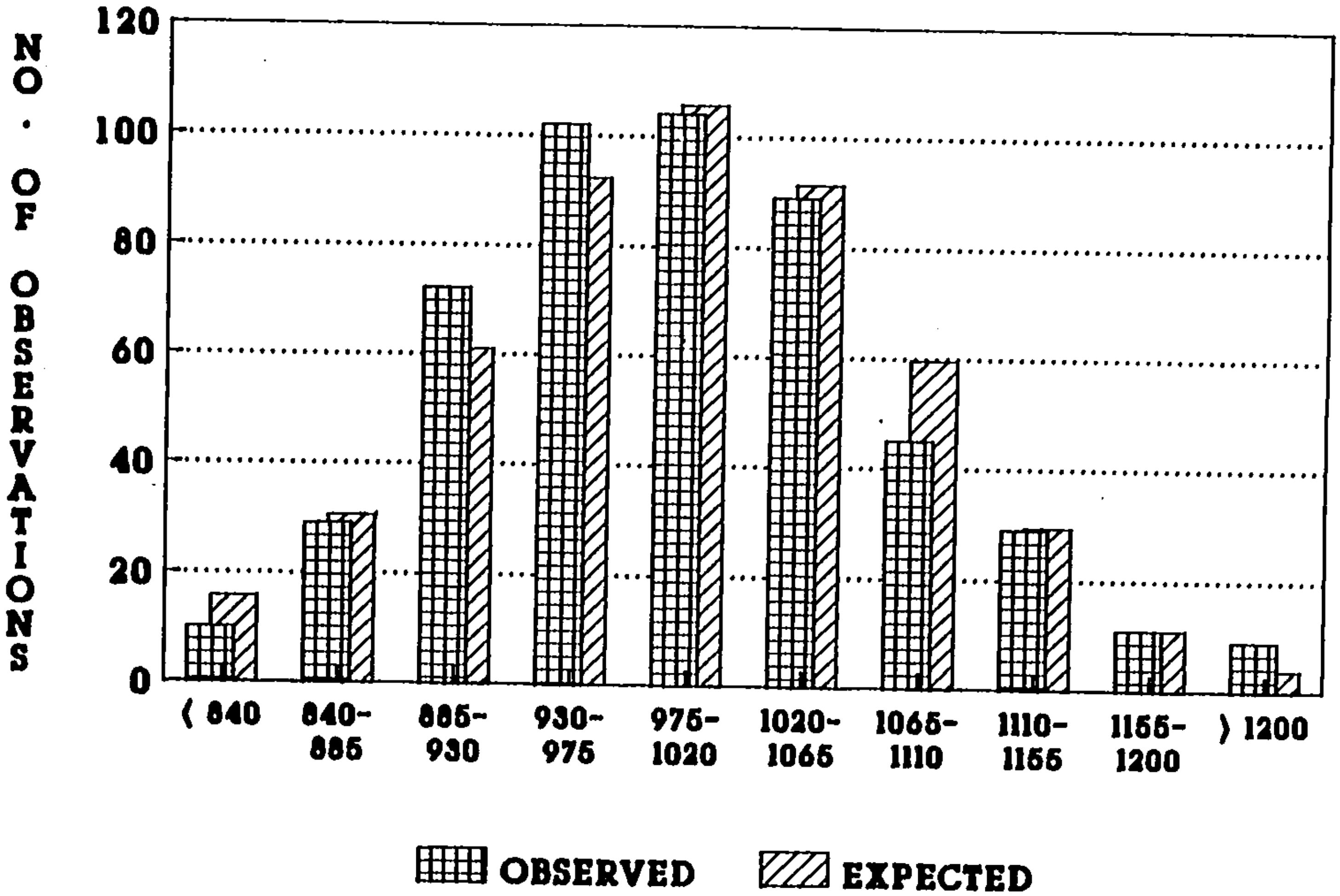


Figure 4.15: Frequency Histograms of the 500 100-Year Design Floods Estimated by Type I Censoring for Data Set 2 (Table 4.3) and the Expected Number if Normally Distributed

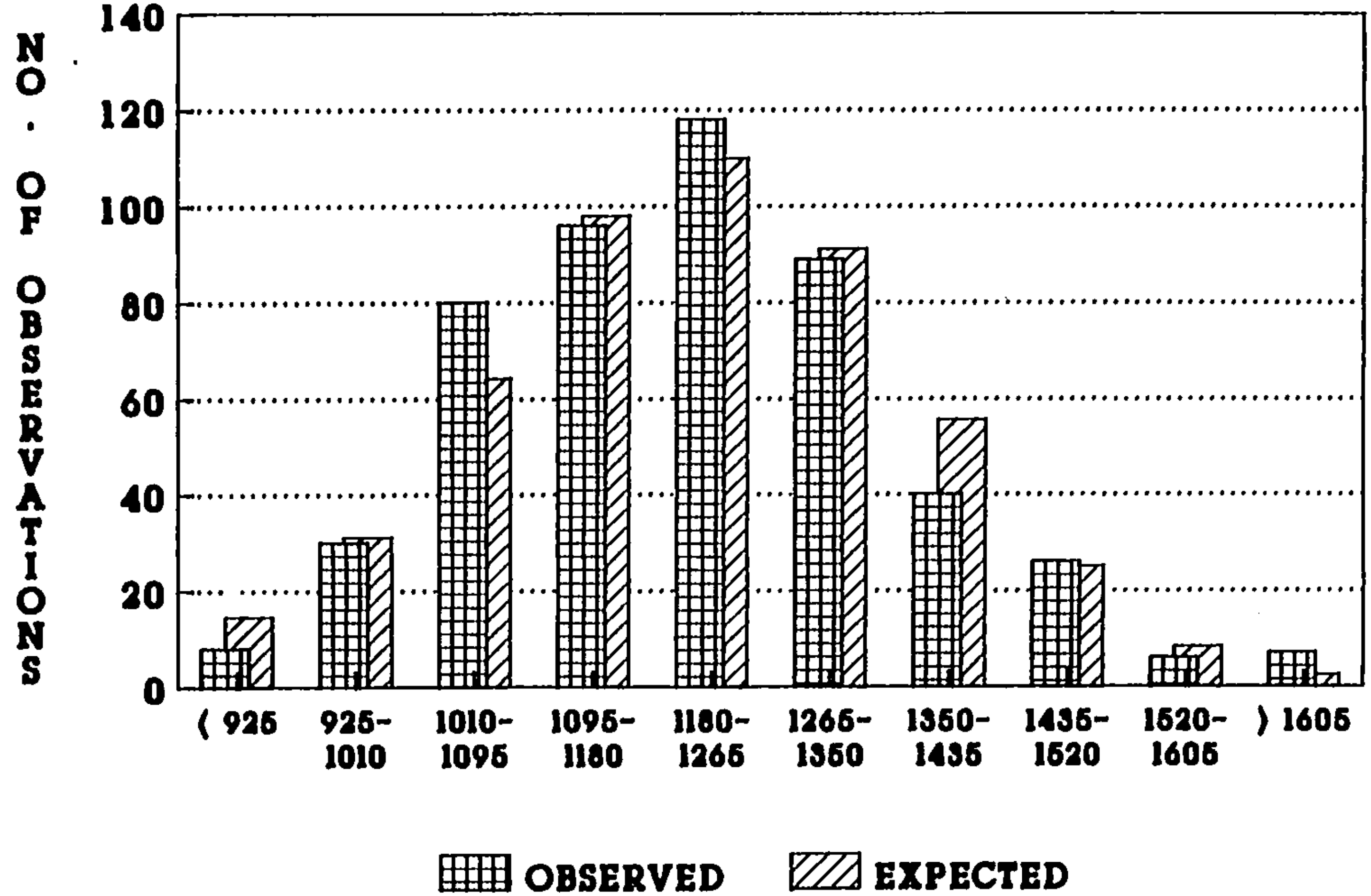


Figure 4.16: Frequency Histograms of the 500 500-Year Design Floods Estimated by Type I Censoring for Data Set 2 (Table 4.3) and the Expected Number if Normally Distributed

Table 4.14 Statistics of the 500 Estimated Design Floods for the Type I Case

Statistic	Design Flood Q_{10}	Q_{50}	Q_{100}	Q_{500}
Data Set 1				
Mean	353.4	420.5	444.9	495.7
Standard Deviation	12.9	21.9	28.2	47.2
Coefficient of Skew	.052	.268	.435	.894
Kurtosis	2.79	2.77	3.10	4.67
Coefficient of Excess	-.21	-.23	0.10	1.67
Data Set 2				
Mean	717.3	910.4	996.6	1,212.4
Standard Deviation	29.7	62.5	84.0	151.9
Coefficient of Skew	.094	.374	.417	.505
Kurtosis	3.02	3.41	3.48	3.68
Coefficient of Excess	0.02	0.41	0.48	0.68

Table 4.15 Chi-squared Test for Goodness of Fit of Quantile Estimates of Type I Censoring to Normal

	Q_{10}	Q_{50}	Q_{100}	Q_{500}
Type I Censoring				
Data Set 1	10.00(ν=6)[1]	13.29(ν=7)	26.39(ν=6)[2]	40.93(ν=7)[2]
Data Set 2	3.76(ν=7)	10.89(ν=7)	10.53(ν=6)	12.49(ν=6)

1: ν implies the degrees of freedom for the χ^2 test.
2: computed χ^2 is greater than the tabulated χ^2 at the 5% level of significance

Table 4.15 lists the computed X^2 statistics for each quantile and data set. The X^2 is computed where the expected number of observations corresponds with that resulting from a normal distribution. The computed X^2 exceeds the tabulated value only for the 100- and 500-year floods of the first data set.

Application of the moment ratio and chi-squared tests indicate that the quantile estimates are distinguishable from the normal distribution some of the time. The quantile estimates increasingly deviate from the normality assumption as the quantile's exceedance probability increases.

This discovery is the same as found for Type II censoring. And again its significance should be recognized for practical applications.

4.5 Effect of Measurement Error for the Worst Flood in Memory Case

This portion of the study is restricted to the second data set of Table 4.3. It is felt that such a limitation will not adversely affect the generality of the results, as they would be similar for other parameter commbinations. The addition of error to the historic flood decreases the information content. This decrease should be proportional to the error added. The decrease in the information content regarding the worst flood should be reflected by an increase in the root mean square estimate of the quantiles. Thus, as the multiplicative error term increases one would expect that at a certain point the information content would be negative thus making the analysis less accurate than if the information were included. The important issue is how much error must be involved before the inclusion of historic information adversely affects the accuracy of the quantile estimation.

Two combinations of data and historic information availability are investigated to assist in the resonse to the aforementioned question. As previously mentioned, both concern the second parameter combination given in Table 4.3. In the first situation it is assumed that there are fifty observed floods of which one is known to be the worst in 100 years. This combination is identical with that used to verify the accuracy of the asymtomptic standard error of estimate computations of the preceeding section.

Monte Carlo procedures are used to investigate the effect of inaccurately estimating the historic flood. A randomm multiplicative error is applied only to the worst flood in each data series so as to mimic this effect. The theoretical censoring threshold, which is also the value of the worst flood, can be evaluated as 996.

The magnitude of the error is allowed to vary and is preselected for each experiment. A random multiplicative error of $\pm 25\%$ infers that the worst flood in each sequence is altered by a factor t(S) where t is a randomly generated standard normal deviate and S represents the standard deviation of the error expression and is computed as the multiplicative error times the theoretical worst flood divided by 1.96. For an error of $\pm 25\%$, S is .25(996)/1.96 or 127. This implies that a random, normally distributed error is added to the worst flood of each series of the experiment. The error which is added to each serie's worst flood is approximately $\pm 25\%$ of the magnitude of the theoretical worst flood 95% of the time and is assumed to be approximately normally distributed.

Alteration to the largest value of a series may result in it no longer being the largest. As the multiplicative error increases, this will occur more frequently. It can also be shown that the largest value may sometimes become negative in sign following the adjustment for error. When the adjusted value is less than the second largest flood, it is assumed that the second largest value is the worst flood and the experiment is continued. If a negative value is encountered, the experiment is stopped as the magnitude of error is becoming quite unrealistic.

Table 4.16 lists the bias in quantile estimation for the first combination of data and historic information. It is evident that the bias of the quantile estimate increases with an increase in the multiplicative error. Table 4.17 lists the root mean square (RMS) error of the quantile estimate. It is as well evident that the RMS error increases with an increase in the multiplicative error.

Figure 4.17 shows that the RMS error for estimating the 100-year flood varies with the magnitude of the multiplicative error. The discrete values of Figure 4.17 are taken from Table 4.17. These values are joined by a straight line only to visually assist in representing the relationship. The exact location of the line would require extensive Monte Carlo simulation; however, for the purposes of this study the relationship demonstrated by Figure 4.17 is sufficiently precise.

Table 4.16: Monte Carlo Simulation Results for Bias of Quantile Estimation for the Second Data Set when $n_a=1$, $n_b=49$, and $n_c=50$

Design Flood	Multiplicative Error			
Return Period	0%	5%	10%	15%
10	-.05[1]	-1.16	-2.05	-2.78
50	.57	-1.92	-4.11	-5.77
100	.95	-2.14	-4.88	-6.94
500	2.28	-2.23	-6.27	-9.25

1: bias in % of theoretical

Table 4.17: Monte Carlo Simulation Results for Root Mean Square Error of Quantile Estimation for the Second Data Set when $n_a=1$, $n_b=49$, and $n_c=50$

Design Flood	Multiplicative Error			
Return Period	0%	5%	10%	15%
10	36.92	37.93	39.87	42.33
50	75.38	78.28	86.30	93.34
100	99.53	103.0	113.9	122.9
500	176.2	178.3	194.9	207.7

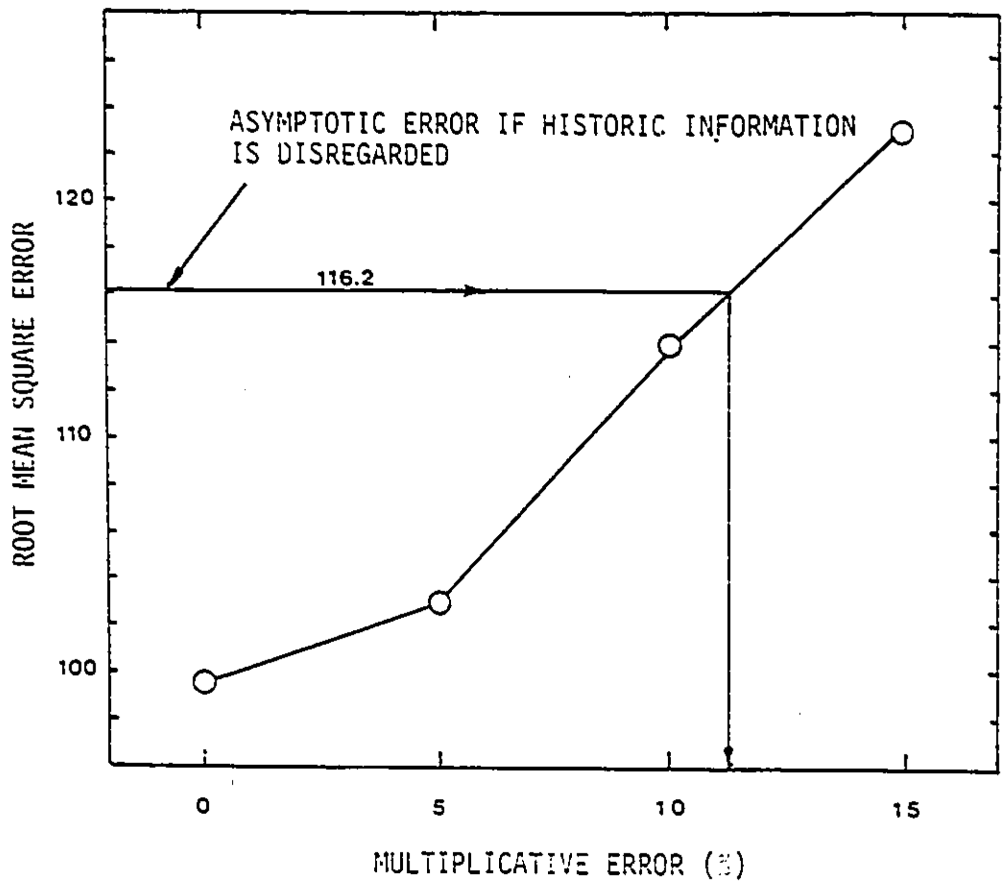


Figure 4.17: Monte Carlo Simulation Results for Root Mean Square Error for the 100-Year Flood when $n_a=1$, $n_b=49$, and $n_c=50$.

If an historic flood had not occurred, there would only be 49 observations upon which to base an estimate of the 100-year or other design floods. This would be the equivalent data if both the historic flood and the historic information are excluded from the analysis. The resultant asymptotic standard-error of estimate of the 100-year flood is 116.2. This value can be obtained using the program listed in Appendix A.

The asymptotic error is displayed on Figure 4.17. When the multiplicative error exceeds approximately 11% the inclusion of historic information decreases the accuracy of the analysis more than if it had been discarded. Given that 49 years of systematic record exists, it is plausible that the worst flood in 100 years would not be that different from the data used to construct the rating curve. Thus, the degree of extrapolation would decrease as the systematic record increases and the number of rating points increase to cover the range of the discharge-rating curve. From such a combination of data and information, it is possible that the historic flood would be estimated with the same order of magnitude of error as the systematically recorded portion of the data. Thus, if the error is less than 11% the historic information contributes to the assessment of flood quantiles. Improved accuracies in the estimation of the historic flood translate to significant improvements in the estimation of flood quantiles, as is evident from Figure 4.17.

The second combination of data and historic information availability assumes that there are fifteen observed floods of which one is known to be the worst in 100 years. This situation represents a decrease in systematic data and an increase in historic information as compared with the first situation.

Tables 4.18 and 4.19 list the bias and RMS error for the 10, 50, 100, and 500-year flood estimates for various multiplicative errors. Again, it is evident that bias and RMS error increase with an increase in the multiplicative error. Figure 4.18 shows the computed RMS error for each multiplicative error as listed in Table 4.19. The known points are connected by straight lines for interpretation purposes only.

Included in Figure 4.18 is a line drawn at an RMS error of 217.3. This value corresponds to the asymptotic standard error of estimate for the given parameter combination if all of the historic information is disregarded. That is, it is assumed that there are only fourteen observations upon which to base an estimate of the 100-year flood.

When a multiplicative error of 50% was attempted, it resulted in the creation of negative values of the worst flood in 100 years. Thus, the experiment was terminated at 40%. Figure 4.18 shows that the RMS error decreased between a multiplicative error of 30 and 40 percent. It is expected that the RMS error should increase as the multiplicative error increases. The lower than anticipated results for the 40% results indicates that an upper limit for the alteration of the worst flood by the approach used herein was reached. For the conditions of this experiment, results indicate that historic information contributed greatly to the

Table 4.18: Monte Carlo Simulation Results for Bias of Quantile Estimation for the Second Data Set when n_a=1, n_b=14, and n_c=85

Design Flood Return Period	Multiplicative Error 0%	10%	15%	20%	30%	40%
10	1.49[1]	-2.49	-3.54	-4.62	-6.16	-6.17
50	0.94	-6.69	-8.42	-10.4	-11.6	-11.6
100	0.84	-8.31	-10.3	-12.6	-13.7	-13.8
500	1.12	-11.5	-14.8	-17.1	-17.9	-18.4

1: bias is in % of theoretical.

Table 4.19: Monte Carlo Simulation Results for Root Mean Square Error of Quantile Estimation for the Second Data Set when n_a=1, n_b=14, n_c=85

Design Flood Return Period	Multiplicative Error 0%	10%	15%	20%	30%	40%
10	52.59	60.26	67.50	69.58	79.14	76.20
50	90.04	115.1	129.8	139.5	153.1	148.7
100	117.1	148.8	166.7	179.6	193.6	189.7
500	207.3	247.4	272.9	292.8	306.3	305.4

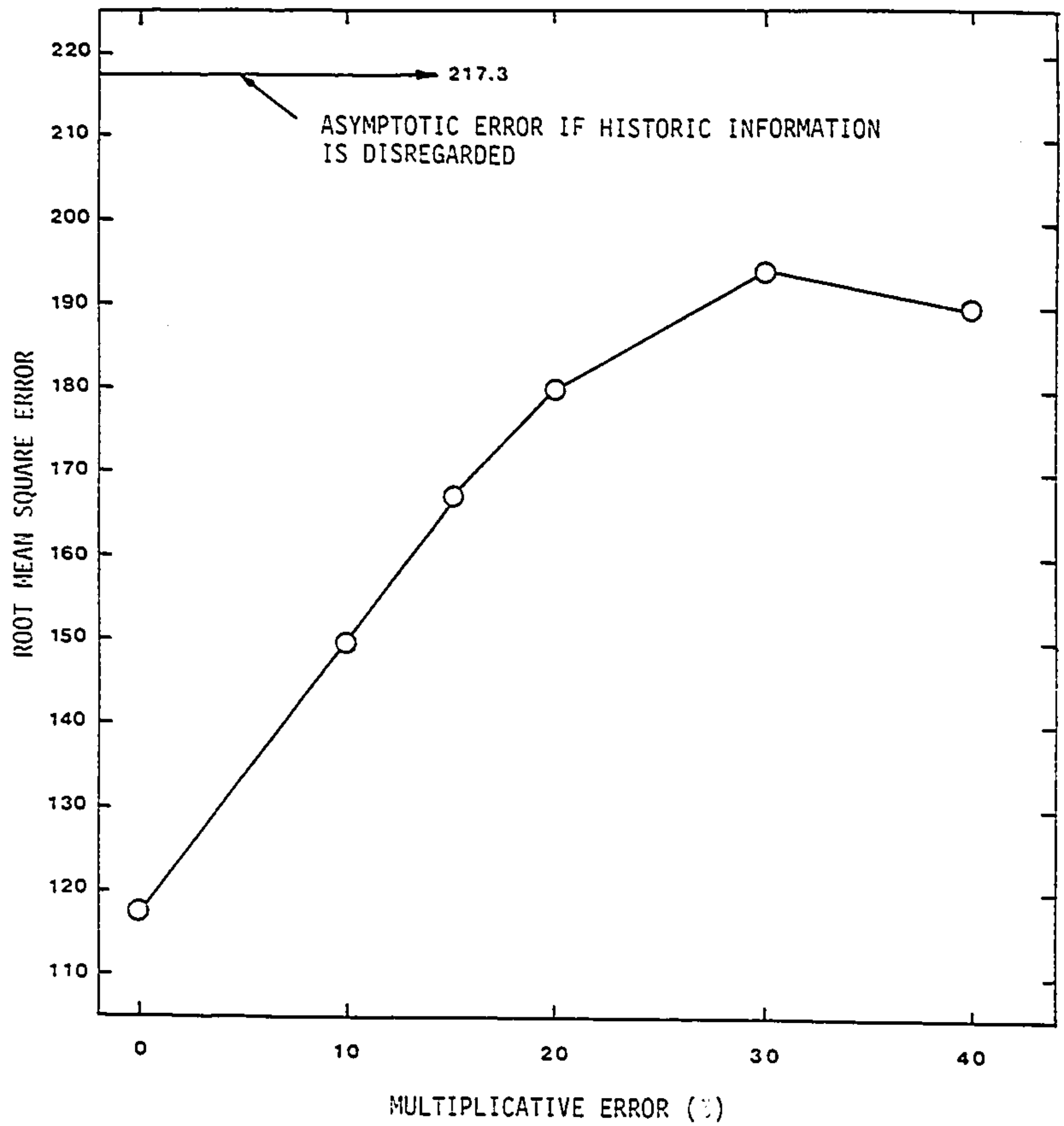


Figure 4.18: Monte Carlo Simulation Results for Root Mean Square Error for the 100-Year Flood when n_a=1, n_b=14, and n_c=85.

accuracy of the estimate of the 100-year flood. This is evident even when the multiplicative error reaches upwards of 30%.

There are only fourteen years of rating points to determine the rating curve as compared with forty-nine for the first situation. It is possible that the error when extrapolating for the historic flood of the second situation would be larger than for the first. Figure 4.18 shows that the historic information is never redundant and increases the accuracy of the quantile's estimate.

Figure 4.19 shows the decrease in the asymptotic standard error of estimate of the 100-year flood as the historic time span increases. For the historic model, it is assumed that there are fifteen observations with one being the worst in memory. The error appears to be asymptotic to approximately 80 for practical purposes. For example, if paleoflood frequency analysis was performed, then the error would not be less than 80. As the historic time span increases, particularly from approximately 100 to 200 years, the reduction in the asymptotic error is negligable. This observation means that efforts of obtaining paleoflood information (i.e. with extremely long historic time spans) might not be warranted. Instead, it seems that the significant increases in accuracy are realized in the 100 to 200 year range of historic time span. The asymptotic standard error of estimate of the 100-year flood when no historic information is known is also presented for comparative purposes. This asymptotic error decreases as the number of observatios, N, increases.

The assessment of the worth of historic information can be facilitated by the use of Figure 4.19. For the second situation where there are fifteen observations and the historic time span is 100 years, it is seen that the asymptotic error is approximately equivalent to that of a conventional analysis with 48 observations. Thus, the knowledge that the largest of a sample of fifteen was the worst in 100 years decreases the asymptotic error equivalent to an additional 33 years of record. This represents a 220% gain in record length.

The above findings have significant implications in hydrometric network design and in frequency analysis.

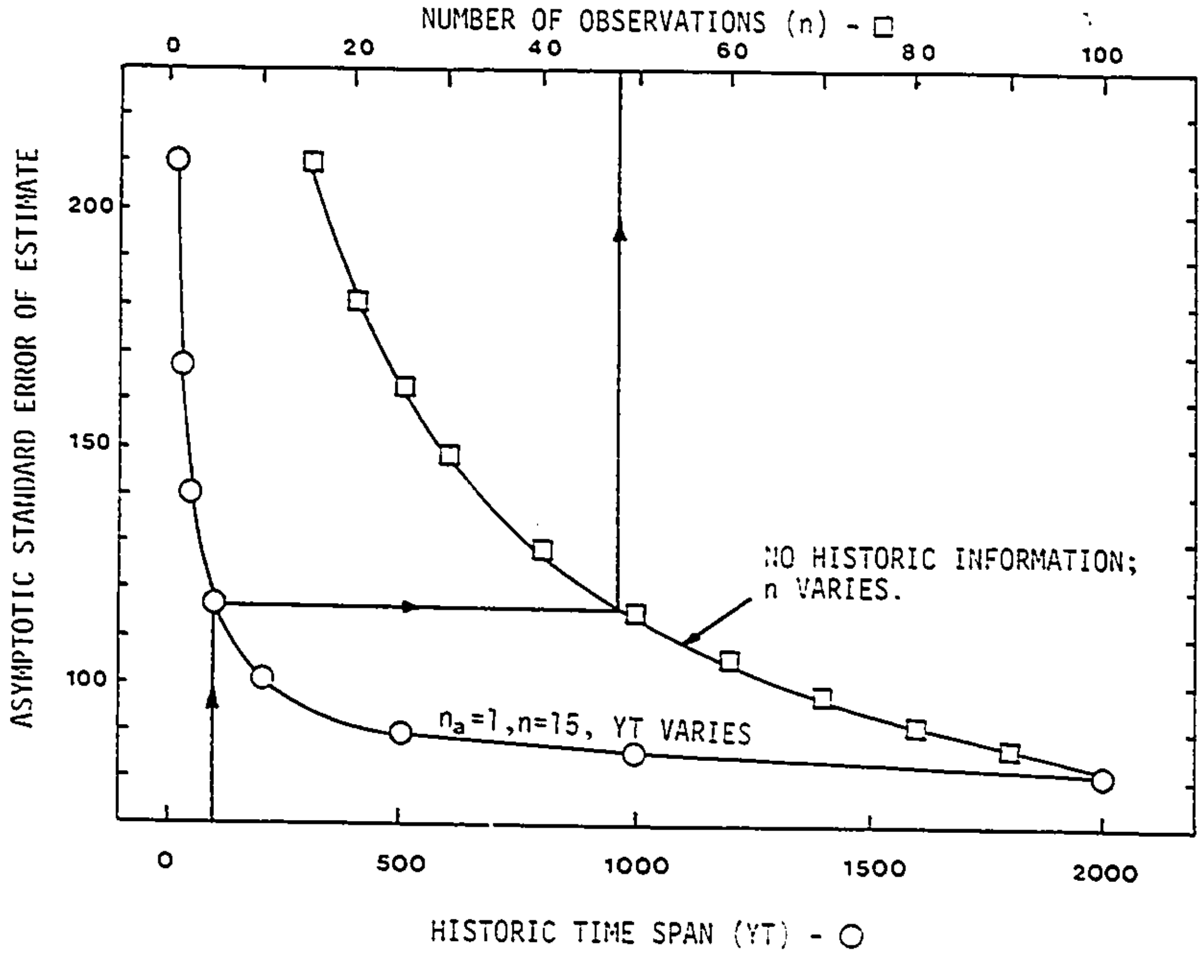


Figure 4.19: Influence of Censored Data on the Asymptotic Standard Error of Estimate of the 100-Year Flood for the Second Data Set.

CHAPTER 5

CASE STUDIES OF FREQUENCY ANALYSIS WITH HISTORIC INFORMATION

The theoretical developments of Chapters 2 and 3 have been incorporated into a computer program called LP3NET. Two case studies of analyzing samples with historic information are presented in this chapter in order to illustrate practical applications. The two hydrometric stations were selected because they are used in the official documentation of frequency analysis techniques in the United States of America (Floyd River at James Iowa) and in Canada (Boyne River near Carman).

The computer program, LP3NET, listed and documented in Appendix B is used to perform the analysis. This program has been developed in the guise of the Consolidated Frequency Analysis (CFA) Package (Pilon et al 1985a). That is, samples with or without historic information and samples containing zeros and/or low outliers can be analyzed.

5.1 Analysis of the Boyne River near Carman - 05OF003

An historical flood frequency analysis of the data for this site is presented by Pilon et al. (1985a). These data are used herein as it is an interesting example and represents an excellent demonstration of the application of frequency analysis when historic information is present.

This gauging station was established in 1919 and remains active today. The data are collected, reviewed, archived, and published by the

Water Resources Branch of Environment Canada. The annual floods for the years 1920, 1921, 1923 through 1926, 1929, and 1931 through 1955 are missing. Data and information up to and including 1982 are used so that results are comparable with those published by Pilon et al. (1985a).

A flood of 187 m^3/s occurred in 1893 and is the largest event ever witnessed at the site. The fourth highest flood occurred in 1970 and had a magnitude of 105 m^3/s. This event is known to be the worst event to have occurred since the flood of 1893. Thus, the censoring threshold is set at 105 m^3/s. Four floods are equal to or larger than this value. Table 5.1 lists the data available at the site. From this table, the total time span of the analysis is seen to be from 1893 to 1982 and represents 90 years. The number of observed peaks is 33 with 29 of these values being less than the censoring threshold. A total of 57 censored peaks are known to have occurred with a magnitude less the censoring threshold of 105 m^3/s.

Figure 5.1 shows the cumulative density function of the LP3 distribution. Parameters are etimated using the maximum likelihood and censoring approach as developed in Chapter 2 and are obtained using the program LP3NET. The parameters of the LP3 are "a" = -.240, b = 17.07, and m = 7.153. Table 5.2 lists design floods corresponding with various probabilities of exceedance for this parameter combination. As parameter "a" is negative, the distribution is upper bounded at exp(m) which corresponds to 1,277 m^3/s. This boundary does not adversely influence the magnitude of the design floods of Table 5.2.

Table 5.1 Annual Flood Peaks: Boyne River near Carman, 05OF003

Year	Flood (m^3/s)	Year	Flood (m^3/s)
1893	187.	1967	37.9
1919	13.5	1968	35.7
1922	15.3	1969	69.7
1927	29.7	1970	105.
1928	12.6	1971	54.1
1930	34.0	1972	19.3
1956	56.1	1973	1.18
1957	10.8	1974	132.
1958	6.09	1975	11.4
1959	14.9	1976	34.8
1960	43.0	1977	7.39
1961	19.4	1978	23.8
1962	38.8	1979	119.
1963	14.5	1980	10.7
1964	13.9	1981	5.47
1965	59.5	1982	6.53
1966	55.2		

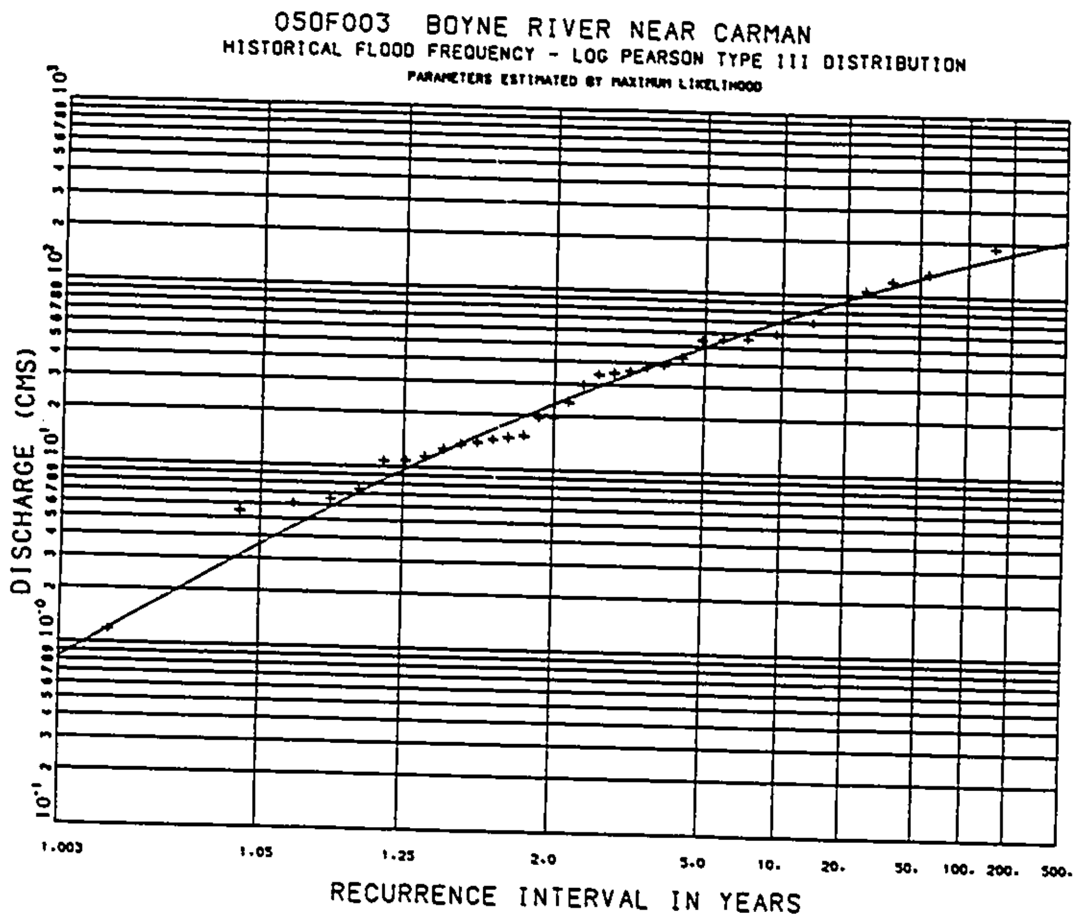


Figure 5.1: Flood Frequency Analysis: Boyne River near Carman, 050F003

Table 5.2 Tabular Flood Frequency Regime for the Boyne River near Carman, 05OF003

Return Period (Years)	Exceedance Probability	Design Flood (m^3/s)
2	.500	23.0
5	.200	49.5
10	.100	71.0
20	.050	93.6
50	.020	125.
100	.010	150.
200	.005	175.
500	.002	209.

Figure 5.1 shows that the data appear to follow the theoretical LP3 curve. The existence of an upper boundary does not appear to influence the shape of the curve with regards to the data points for the plotted range of the exceedance probabilities. The graph represents a logarithmic probability plot of the data. The plotting positions are due to Cunnane (1978) with the rank adjustment for historic information as suggested by Benson (1950).

For comparison, an analysis is performed on the same data with procedures for the identification and treatment of low outliers. These procedures are as outlined by the Hydrology Subcommittee (1982) and as have been incorporated into the CFA Package (Pilon et al. 1985a). Application of the procedure identifies the flood of 1973, 1.18 m^3/s, as being a low outlier. LP3NET adopts the procedures of using "synthetic statistics" for fitting the distribution when a zero or low outlier is identified as outlined by the Hydrology Subcommittee (1982) and Pilon et al. (1985a).

Figure 5.2 shows the cumulative density function of the LP3 distribution which has been adjusted due to the identification of a low outlier. The adjusted parameters "a", b, and m of the LP3 distribution are found to be 0.03316, 632.4, and -17.86, respectively. Table 5.3 gives the tabular frequency regime for various exceedance probabilities corresponding to this parameter combination.

A comparison of Figure 5.1 with Figure 5.2 shows the treatment of the low outlier results in an upwarding shifting of the cumulative density function above the 10-year flood level and a downward shift below this

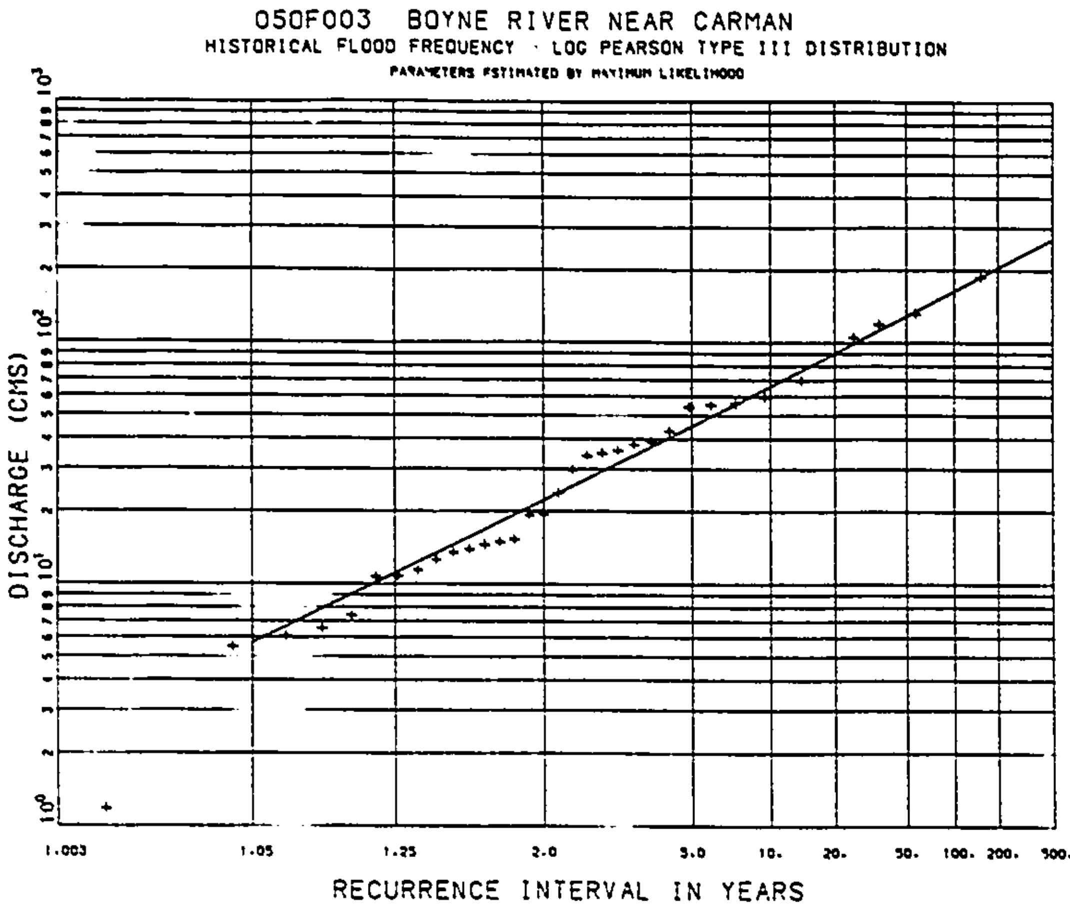


Figure 5.2: Flood Frequency Analysis Adjusted for One Low Outlier: Boyne River near Carman, 050F003

Table 5.3 Tabular Flood Frequency Regime for the Boyne River near Carman, 05OF003, Adjusted for One Low Outlier

Return Period (Years)	Exceedance Probability	Design Flood (m^3/s)
2	.500	22.2
5	.200	45.1
10	.100	65.8
20	.050	90.2
50	.020	129.
100	.010	164.
200	.005	205.
500	.002	269.

value. Tables 5.2 and 5.3 permit a numeric comparison of the treatment process. Treatment for the presence of the low outlier results in an increase of about 9.3% in the estimate of the 100-year flood and a decrease of about 8.9% in the estimate of the 5-year flood. The estimate of the 500-year flood changes by approximately 29%.

The identification of low outliers and the subsequent adjustment of the frequency curve due to their identification can have substantial influences on the estimate of the design flood. A visual comparison of Figure 5.2 with Figure 5.3 substantiates this claim. It is interesting to note that the data points more closely follow the theoretical frequency curve when it is adjusted for the presence of low outliers. The asymptotic standard error of estimate of the 2-, 10-, and 100-year floods of Table 5.3 are 13.8%, 15.4%, and 21.9%, respectively, based on the theoretical developments of Chapter 3.

5.2 Analysis of the Floyd River at James, Iowa - 06-6005

The national guidelines for flood frequency analysis at gauged sites in the United States (Hydrology Subcommittee 1982) contain several practical examples, some of which are concerned with the inclusion of historic information. The Floyd River at James, Iowa, is used to demonstrate their approaches to frequency analysis when historic information pertaining to the site is documented. The national guidelines of the United States proposes an historically weighted moments method of fitting the LP3 distribution. Note that the data and all results pertaining to this section are given in the English system to facilitate comparison of results with those of the report of the Hydrology Subcommittee (1982).

Table 5.4 lists the data available at the site. The sample contains 39 annual flood peaks with the largest recorded flood occurring in 1953 with a value of 71,500 ft^3/s (2,025 m^3/s). This flood is much larger than the second largest flood which occurred in 1962 with a magnitude of 20,600 ft^3/s (583 m^3/s). "Information from local residents indicates that the 1953 event is known to be the largest event since 1892" (Hydrology Subcommittee 1982, p. 12-17). Thus, historical information is available and the maximum likelihood and censoring approach as developed in Chapter 2 is applied.

Table 5.4 Annual Flood Peaks: Floyd River at James, Iowa 06-6005

Year	Flood (ft^3/s)	Year	Flood (ft^3/s)
1935	1,460	1955	2,260
1936	4,050	1956	318
1937	3,570	1957	1,330
1938	2,060	1958	970
1939	1,300	1959	1,920
1940	1,390	1960	15,100
1941	1,720	1961	2,870
1942	6,280	1962	20,600
1943	1,360	1963	3,810
1944	7,440	1964	726
1945	5,320	1965	7,500
1946	1,400	1966	7,170
1947	3,240	1967	2,000
1948	2,710	1968	829
1949	4,520	1969	17,300
1950	4,840	1970	4,740
1951	8,320	1971	13,400
1952	13,900	1972	2,940
1953	71,500	1973	5,660
1954	6,250		

The censoring threshold is set at 71,500 ft^3/s. The analysis includes data and information from 1892 to 1973, inclusive. Note that neither data nor information from 1974 to the present time (1988) are included so that results are comparable with those of the report of the Hydrology Subcommittee (1982). Hence, the total time span of the analysis is 82 years of which 43 are censored data and 39 are fully defined floods.

Figure 5.3 shows the plot of the observed floods and the cumulative density function of the LP3 distribution. Results are obtained using the program LP3NET. Parameters a, b, and m are estimated to be 0.1239, 69.55, and -.4451, respectively. No low outliers are detected, thus no adjustments to the parameters are made.

Table 5.5 lists the design floods corresponding with various exceedance probabilities for this parameter combination. The distribution is lower bounded at exp(m) or 0.641 ft^3/s and is unbounded above. The asymptotic standard error of estimate for the 2-, 10-, and 100-year floods of Table 5.5 are 17.7%, 22.9%, and 41.5%, respectively.

Th Hydrology Subcommittee (1982) give the historically adjusted mean, standard deviation, and skewness to be 3.5375, 0.4377, and 0.1650 for the site. Note that the mean and standard deviation are derived using common rather than natural logarithms. Table 5.6 lists the design floods corresponding with various exceedance probabilities based on the at-site adjusted moments.

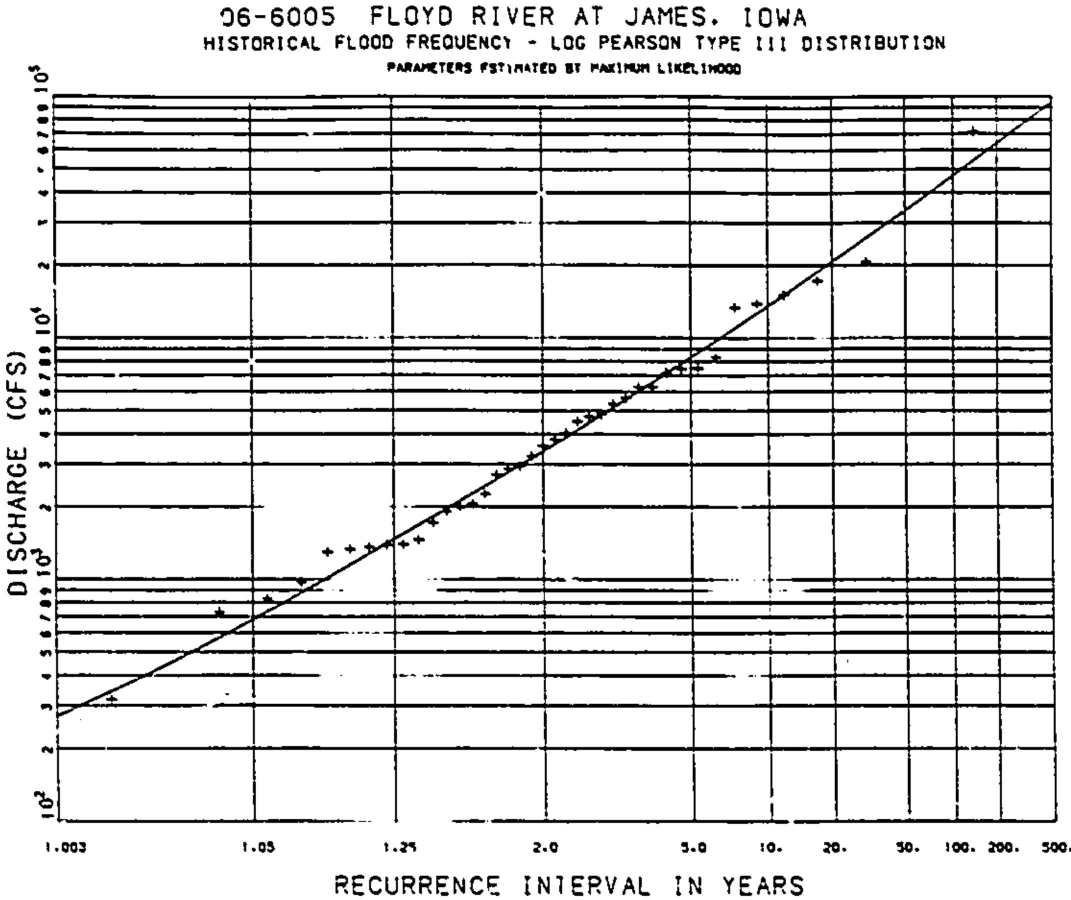


Figure 5.3: Flood Frequency Analysis: Floyd River at James, Iowa 06-6005

Table 5.5 Tabular Flood Frequency Regime for the Floyd River at James, Iowa 06-6005

Return Period (Years)	Exceedance Probability	Design Flood (ft^3/s)
2	.500	3,400
5	.200	8,320
10	.100	13,600
20	.050	20,700
50	.020	33,700
100	.010	47,000
200	.005	64,000
500	.002	93,900

Table 5.6 Tabular Flood Frequency Regime Derived using Historically Adjusted Moments for the Floyd River at James, Iowa 06-6005

Return Period (Years)	Exceedance Probability	Design Flood (ft^3/s)
2	.500	3,540
5	.200	7,980
10	.100	12,800
20	.050	18,900
50	.020	29,800
100	.010	40,600
200	.005	54,000
500	.002	76,800

A comparison of Table 5.5 with Table 5.6 shows that the maximum likelihood approach gives design flood estimates which tend to be higher than the adjusted moments approach for the Floyd River at James, Iowa.

In summary, the developments in this thesis enable the estimation of parameters and the asymptotic standard error of estimate of the quantiles.

As a caution to users, it should be noted that the maximum likelihood approach might occasionally not yield estimates of the parameters. Also, sometimes there might be no solution for the asymptotic error. The lack of a solution, although infrequent, is due to an instability in the Fisher information matrix.

CHAPTER 6

CONCLUSIONS AND RECOMMENDATIONS

6.1 Conclusions

Using maximum likelihood and censored sample theories, equations have been developed to estimate the parameters and the asymptotic standard error of estimate of the parameters and quantiles for the gamma family of distributions. These developments have been incorporated into a computer program for the log Pearson Type III distribution.

A Monte Carlo study for the log Pearson Type III distribution indicated that:

i) quantiles are generally estimated without statistically significant bias;

ii) the asymptotic results for the standard error of the estimated quantiles are accurate; and

iii) design flood estimates for a particular quantile are not normally distributed. This latter finding is of importance if and when confidence limits are derived using the design flood estimates and their asymptotic error.

The Monte Carlo study also indicated that some currently used procedures in the hydrologic literature for gamma variate generation are biased.

Results of the study for the Type I censored data indicated that quantile estimates were sometimes statistically biased. The bias when estimating the 10-, 50-, 100-, and 500-year floods did not exceed 1.5% of the theoretical value. This low degree of bias was not considered practically significant. As with the Type II censored data, the derived theoretical asymptotic standard error of the estimated quantiles were accurate. The design floods estimated for a particular quantile deviate from normality, especially as the quantile's exceedance probability increases.

The study of both the Monte Carlo results and the theoretical expressions for the asymptotic standard error of estimate indicate that the inclusion of historic information can significantly increase the accuracy of the quantile estimation procedures. However, the rate of increase diminishes sharply with an increase in the historic time span beyond approximately 100 years. This observation reflects negatively upon the worth of paleoflood information in flood frequency analysis.

The study also indicates that even when the historic flood is estimated with considerable error, improvements in accuracy are still realized.

The findings of this study have significant implications for flood frequency analysis both from theoretic and practical considerations.

6.2 Recommendations

1) Contrary to several published and widely used procedures,

this study showed that the normality assumption of the quantile's distribution is not valid. Therefore, future research should be directed towards the determination of the distribution of the quantiles' estimates, especially as it pertains to confidence limits.

2) The theoretical developments presented in this thesis could be used to assist in the design of hydrometric networks.

REFERENCES

Abramaowitz, M. and Stegun, I.A. "Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables", U.S. Department of Commerce, National Bureau of Standards, Applied Mathematics Series, 55, Issued June 1964, with corrections December 1972.

Adamowski, K. "Nonparametric Kernal Estimation of Flood Frequencies", Water Resources Research, 21(11), pp. 1585-1590, 1985.

Alexander, W.J.R. "Flood Hydrology - A South African Handbook", draft, Department of Civil Engineering, University of Pretoria, South Africa, 1988.

Ashkar, F. and Bobée, B. "Variance of the T-Year Event in the Log Pearson Type-3 Distribution - Comment", Journal of Hydrology, 84, pp. 181-187, 1986.

Ashkar, F. and Bobée, B. "Estimating Parameters of Flood Frequency Distributions by the Generalized Method of Moments with Special Emphasis on the Log Pearson Type 3 Distribution", Proceedings of the 8th Canadian Hydrotechnical Conference, Montreal, Quebec, May 19-22, 1987.

Ashkar, F. and Bobée, B. "Confidence Intervals for Flood Events Under a Pearson 3 or Log Pearson 3 Distribution", Water Resources Bulletin, 24(3), pp. 639-650, 1988.

Baker, V.R., Kochel, R.C. and Patton, P.C. "Long-term Flood Frequency Analysis using Geological Data", IASH Publication, No. 128, pp. 3-9, 1979.

Baker, V.R. "Paleoflood Hydrologic Techniques for the Extension of Stream Flow Records", Transportation Research Record, 922, pp. 18-23, 1983.

Baker, V.R. Ely, L.L., O'Connor, J.E. and Partridge, J.B. "Paleoflood Hydrology and Design Applications", Regional Flood Frequency Analysis, D. Reidel Publishing Company, Proceedings of the International Symposium on Flood Frequency and Risk Analyses, 14-17 May 1986, Baton Rouge, U.S.A., pp. 339-354, 1986a.

Baker, V.R., Pickup, G. and Webb, R.H. "Paleoflood Hydrologic Analysis at Ungaged Sites, Central and Northern Australia", Regional Flood Frequency Analysis, D. Reidel Publishing Company, Proceedings of the International Symposium on Flood Frequency and Risk Analyses, 14-17 May 1986, Baton Rouge, U.S.A., pp. 325-338, 1986b.

Baker, V.R. "Paleoflood Hydrology and Hydroclimatic Change", IAHS Publication No. 168, pp. 123-131, 1987a.

Baker, V.R. "Paleoflood Hydrology and Extraordinary Flood Events", Journal of Hydrology, 96, pp. 79-99, 1987b.

Benson, M.A. "Use of Historical Data in Flood-Frequency Analysis", Transactions of the American Geophysical Union, 31(3), pp. 419-424, 1950.

Benson, M.A. "Uniform Flood-Frequency Estimating Methods for Federal Agencies", Water Resources Research, 4(5), pp. 891-908, 1968.

Bobée, B. "Sample Error of T-Year Events computed by Fitting a Pearson Type 3 Distribution", Water Resources Research, 9(5), pp. 1264-1270, 1973.

Bobée, B. "The Log Pearson Type 3 Distribution and Its Application in Hydrology", Water Resources Research, 11(5), pp. 681-689, 1975.

Bobée, B. and Robitaille, R. "Correction of Bias in the Estimation of the Coefficient of Skewness", Water Resources Research, 11(6), pp. 851-854, 1975.

Bobée, B. and Robitaille, R. "The Use of the Pearson Type 3 and Log Pearson Type 3 Distributions Revisited", Water Resources Research, 13(2), pp. 427-443, 1977.

Bobée, B. "Comment on 'Fitting the Pearson Type 3 Distribution in Practice' by J. Buckett and F.R. Oliver", Water Resources Research, 15(3), p. 730, 1979.

Buckett, J. and Oliver, F.R. "Fitting the Pearson Type 3 Distribution in Practice", Water Resources Research, 13(5), pp. 851-852, 1977.

Chen, C., Yeh, Y. and Tan, W. "The Important Role of Historical Flood Data in the Estimation of Spillway Design Floods", Scientia Sinica, 18(5), pp. 669-680, 1974.

Cheng, R.C. and Feast, G.M. "Gamma Variate Generators with Increased Shape Parameter Range", Communications of the ACM, 23, pp. 389-393, 1980.

Cohen, A.C. "Estimating the Mean and Variance of Normal Populations for Singly Truncated and Doubly Truncated Samples", Annals of Mathematical Statistics, 21(4), pp. 557-69, 1950.

Cohn, T.A. "The Incorporation of Historical Information in Flood Frequency Analysis", M.S. Thesis, Cornell University, 79 pp., August, 1984.

Cohn, T.A. "Flood Frequency Analysis with Historical Flood Information", Ph.D. Dissertation, Cornell University, 144 pp., June, 1986.

Cohn, T.A. and Stedinger, J.R. "Use of Historical Information in a Maximum-Likelihood Framework", Journal of Hydrology, 96, pp. 215-224, 1987.

Condie, R. "The Log Pearson Type 3 Distribution: The T-Year Event and Its Asymptotic Standard Error by Maximum Likelihood Theory", Water Resources Research, 13(6), pp. 987-991, 1977.

Condie, R. "Frequency Analysis of Negatively Skewed Runoff Samples", Proceedings of the Third CSCE Atlantic Regional Hydrotechnical Conference, Halifax, N.S., February 13-14, 1978.

Condie, R., Nix, G.A., and Boone, L.G. "Flood Damage Reduction Program Flood Frequency Analysis - Program FDRPFFA", Environment Canada, Inland Waters Directorate, 1981.

Condie, R. and Lee, K.A. "Flood Frequency Analysis with Historic Information", Journal of Hydrology, 58, pp. 47-61, 1982.

Condie, R. and Pilon, P.J. "Fitting the Log Pearson Type 3 Distribution to Flood Frequency Analysis with Historic Information", Proceedings of the 6th Canadian Hydrotechnical Conference, June 2-3, Ottawa, Ontario, pp. 11-22, 1983.

Condie, R. "Flood Samples from a Three-Parameter Lognormal Population with Historic Information: The Asymptotic Standard Error of Estimate of the T-Year Flood", Journal of Hydrology, 85, pp. 139-150, 1986.

Costa, J.E. "Holocene Stratigraphy in Flood Frequency Analysis", Water Resources Research, 14(4), pp. 626-632, 1978.

Costa, J.E. "Fluvial Paleoflood Hydrology", Abstract in EOS; Transactions of the American Geophysical Union, 65(45), p. 890, November 6, 1984.

Croley, T.E., II. "Gamma Synthetic Hydrographs", Journal of Hydrology, 47, pp. 41-52, 1980.

Cruise, J.F. and Contractor, D.N. "Unit Hydrographs for Urbanizing Watersheds", ASCE, Journal of the Hydraulics Division, HY3, pp. 440-445, 1980.

Cunnane, C. "Unbiased Plotting Positions - A Review", Journal of Hydrology, 37,pp. 205-222, 1978.

Cunnane, C. "Review of Statistical Models for Flood Frequency Estimation", Hydrologic Frequency Modeling, D. Riedel Publishing Company, Proceedings of the International Symposium on Flood Frequency and Risk Analyses, 14-17 May 1986, Baton Rouge, U.S.A., pp. 49-96, 1987.

Des Groseilliers, L. "Algorithms recommandés pour la génération de variates gamma", INRS-Eau, Rapport interne no. 100, University du Québec, 1985.

DEC, "Programming in VAX FORTRAN", Appendix D FORTRAN Language Summary, Digital Equipment Corporation, Maynard, Mass., p. D-42, 1984.

Fisher, R.A. "On an Absolute Criterion for Fitting Frequency Curves", Messenger of Mathematics, xli, pp. 155-160, 1912.

Fisher, R.A. "On the Mathematical Foundations of Theoretical Statistics", Philosophical Transactions of the Royal Society of London, Series A, 222, pp. 309-368, 1922.

Gueven, C. "A Simplified Semiempirical Approach to Probabilities of Extreme Hydrologic Droughts", Water Resources Research, 9(2), pp. 441-453, 1983.

Hahn, G.J. and Shapiro, S.S. "Statistical Models in Engineering", John Wiley & Sons, New York, 1967.

Hald, A. "Maximum Likelihood Estimation of the Parameters of a Normal Distribution which is Truncated at a Known Point", Skandinavisk Aktuarietidskrift, 32(1/2), pp. 119-134, 1949.

Harper, H.L. and Moore, A.H. "Maximum-Likelihood Estimation, from Doubly Censored Samples, of the Parameters of the First Asymptotic Distribution of Extreme Values", American Statistical Association Journal, 63, pp. 889-901, 1968.

Hastings, H. "Approximations for Digital Computers", Princeton University Press, Princeton, N.J., 1955.

Helley, E.J. and Lamarche, V.C. "Historic Flood Information for Northern California Streams from Geologic and Botanical Evidence", U.S. Geological Survey Professional Paper 485-E, 1973.

Hoshi, K. and Burges, S.J. "Sampling Properties of Parameter Estimates for the Log Pearson Type 3 Distribution, using Moments in Real Space", Journal of Hydrology, 53, pp. 305-316, 1981.

Hosking, J.R.M. and Wallis, J.R. "Paleoflood Hydrology and Flood Frequency Analysis", Water Resources Research, 22(4), pp. 543-550, 1986a.

Hosking, J.R.M. and Wallis, J.R. "The Value of Historical Data in Flood Frequency Analysis", Water Resources Research, 22(11), pp. 1606-1612, 1986b.

Hupp, C.R. "Botanical Evidence of Floods and Paleoflood History", Regional Flood Frequency Analysis, D. Reidel Publishing Company, Proceedings of the International Symposium on Flood Frequency and Risk Analyses, 14-17 May 1986, Baton Rouge, U.S.A., pp. 355-370, 1986.

Hydrology Subcommittee. "Guidelines for Determining Flood Flow Frequency", Interagency Advisory Committee on Water Data, U.S. Geological Survey, Office of Water Data Coordination, Reston, Virginia 22092, Bulletin #17B, Revised with Editorial Corrections, 1982.

Kendall, G.R. "Probability Distribution of a Single Variable", Proceedings of Hydrology Symposium No. 5, McGill University, Montreal, February 24-25, 1966.

Kendall, M.G. and Stuart, A. "The Advanced Theory of Statistics", Volume 2, 4th edition, Hafner, New Yrok, 1979.

Kirby, W. "Computer-Oriented Wilson-Hilferty Transformation that Preserves the First Three Moments and the Lower Bound of the Pearson Type 3 Distribution", Water Resources Research, 8(5), pp. 1251-1254, 1972.

Kirby, W. "Algebriac Boundedness of Sample Statistics", Water Resources Research, 10(2), pp. 220-222, 1974.

Kirby, W. "Statistical Error Analysis of the Slope-Area Method of Indirect Discharge Determination", EOS, 66, p. 267, 1985.

Kirby, W. "Linear Error Analysis of Slope-Area Discharge Determinations", Journal of Hydrology, 96, pp. 125-138, 1987.

Kite, G.W. "Confidence Limits for Design Events", Water Resources Research, 11(1), pp. 48-53, 1975.

Kochel, R.C. and Baker, V.R. "Paleoflood Hydrology", Science, 215(4531), pp. 353-361, 22 January 1982.

Landwehr, J.M., Matalas, N.C. and Wallis, J.R. "Some Comparisons of Flood Statistics in Real and Log Space", Water Resources Research, 14(5), pp. 902-920, 1978.

Lane, W.L. "Paleohydrologic Data and Flood Frequency Estimation", Regional Flood Frequency Analysis, D. Reidel Publishing Company, Proceedings of the International Symposium on Flood Frequency and Risk Analyses, 14-17 May 1986, Baton Rouge, U.S.A., pp. 287-298, 1986.

Lawless, J.F. "Statistical Models and Methods for Lifetime Data", John Wiley & Sons, Toronto, 1982.

Leese, M.N. "Use of Censored Data in the Estimation of Gumbel Distribution Parameters for Annual Maximum Flood Series", Water Resources Research, 9(6), pp. 1534-1542, 1973a.

Leese, M.N. "The Use of Censored Data in Estimating T-Year Floods", Proceedings of the UNESCO-WMO-IAHS Symposium, Design of Water Resources Projects with Inadequate Data, Volume 1, pp. 225-247, Madrid, Spain, 1973b.

Liebscher, H.J. "Paleohydrologic Studies using Proxy Data and Observations", IASH Publication No. 168, pp. 111-122, 1987.

Lowery, M.D. and Nash, J.E. "A Comparison of Methods of Fitting the Double Exponential Distribution", Journal of Hydrology, 10, pp. 259-275, 1970.

Markovic, R.D. "Probability Functions of Best Fit to Distributions of Annual Precipitation and Runoff", Colorado State University Hydrology Paper No. 8, August 1965, Fort Collins, Colorado.

Matalas, N.C. and Wallis, J.R. "Eureka! It Fits a Pearson Type 3 Distribution", Water Resources Research, 9(2), pp. 281-289, 1973.

Matalas, N.C., Slack, J.R., and Wallis, J.R. "Regional Skew in Search of a Parent", Water Resources Research, 11(6), pp. 815-826, 1975.

Moss, M.E., Thomas, W.O. Jr., and Scott, A.G. "Cost-Effectiveness of the U.S. Geological Survey Stream-Gauging Program", Transportation Research Record 1017, pp. 8-15, 1985.

Natural Environmental Research Council. "Flood Studies Report Volume I Hydrological Studies", 550 p., London, England, 1975.

Nozdryn-Plotnicki, M.J. and Watt, W.E. "Assessment of Fitting Techniques for the Log Pearson Type 3 Distribution Using Monte Carlo Simulation", Water Resources Research, 15(3), pp. 714-718, 1979.

Pearson, K. "Contributions to the Mathematical Theory of Evolution. II. Skew Variation in Homogeneous Material", Philosophical Transactions of the Royal Society of London, Series A, 186, pp. 343-414, 1895.

Phien, H.N. and Hira, M.A. "Log Pearson Type-3 Distribution: Parameter Estimation", Journal of Hydrology, 64, pp. 25-37, 1983.

Phien, H.N. and Hsu, L.-C. "Variance of the T-Year Event in the Log Pearson Type-3 Distribution", Journal of Hydrology, 77, pp. 141-158, 1985.

Phien, H.N., Sunchindah, A. and Patnaik, D. "Normality of Hydrologic Data", Water Resources Bulletin, 18(1), pp. 37-42, 1982.

Pilon, P.J., Condie, R., and Harvey, K.D. "Consolidated Frequency Analysis Package (CFA) User Manual for Version 1--DEC PRO Series", Environment Canada, Inland Waters Directorate, Ottawa, Ontario, 1985a.

Pilon, P.J., Condie, R., and Harvey, K.D. "Flood Frequency Analysis with Zero Flows, Low Outliers, and Historic Information", Proceedings of the 7th Canadian Hydrotechnical Conference, pp. 253-268, Saskatoon, Saskatchewan, May 27 - June 1, 1985b.

Pilon P.J., Harvey, K.D., and Condie, R. "An Assessment of Contemporary Flood Frequency Distributions", Proceedings of the 8th Canadian Hydrotechnical Conference, Montreal, Quebec, May 19-22, 1987.

Potter, H.R. "The Use of Historic Records for the Augmentation of Hydrological Data", Report No. 46, Institute of Hydrology, Wallingford, Oxon, 59 p., January 1978.

Rao, D.V. "Log Pearson Type 3 Distribution: Method of Mixed Moments", ASCE, Journal of the Hydraulics Division, HY6, pp. 999-1019, 1980.

Shi, F., Yi, Y., and Han, M. "Investigation and Verification of Extraordinarily Large Floods on the Yellow River", Journal of Hydrology, 96, pp. 69-78, 1987.

Sigafoos, R.S. "Botanic Evidence of Floods and Flood Plain Development", U.S. Geological Survey Professional Paper 485-A, 35 p., 1964.

Singh, V.P. and Singh, K. "Derivation of the Pearson Type (PT) III Distribution by Using the Principle of Maximum Entropy (POME)", Journal of Hydrology, 80, pp. 197-214, 1985.

Srikanthan, R. and McMahon, T.A. "Log Pearson III Distribution - Effect of Dependence, Distribution Parameters and Sample Size on Peak Annual Flood Estimates", Journal of Hydrology, 52, pp. 149-159, 1981.

Stacey, E.W. "A Generalization of the Gamma Distribution", The Annals of Mathematical Statistics, 33(3), pp. 1187-1192, 1962.

Stacey, E.W. and Mihram, G.A. "Parameter Estimation for a Generalized Gamma Distribution", Technometrics, 7(3), pp. 349-358, 1965.

Stedinger, J.R., and Cohn, T.A. "Flood Frequency Analysis with Historical and Paleoflood Information", Water Resources Research, 22(5), pp. 785-793, 1986a.

Stedinger, J.R. and Cohn, T.A. "Historical Flood Frequency Data: Its Value and Use", Regional Flood Frequency Analysis, D. Reidel Publishing Company, Proceedings of the International Symposium on Flood Frequency and Risk Analyses, 14-17 May 1986, Baton Rouge, U.S.A., pp. 273-386, 1986b.

Tadikamalla, P.R. and Johnson, M.E. "A Complete Guide to Gamma Variate Generation", American Journal of Mathematical and Management Sciences, 1(3), pp. 213-236, 1981.

Thom, H.C.S. "A Note on the Gamma Distribution", Monthly Weather Review, 86(4), Department of Commerce, pp. 117-122, 1958.

Thomas, W.O. Jr. "A Uniform Technique for Flood Frequency Analysis", Journal of Water Resources Planning and Management,, ASCE, 111(3), pp. 321-337, 1985.

Thomas, W.O. Jr., and Cheng, L. "Hydrometric Network Analysis in the Oldman River Basin, Alberta using Generalized Least Squares", Proceedings of the 7th Hydrotechnical Conference, pp. 283-302, Saskatoon, Saskatchewan, May 27 - June 1, 1985.

Thomson, M.T., Gannon, W.B., Thomas, M.P., and Hayes, G.S. "Historical Floods in New England", U.S. Geological Survey Water Supply Paper 1779-M, 105 p., 1964.

Wallis, J.R., Matalas, N.C., and Slack, J.R. "Just a Moment!", Water Resources Research, 10(2), pp. 211-219, 1974.

Wallis, J.R. and Wood, E.F. "Relative Accuracy of Log Pearson III Procedures", Journal of Hydraulic Engineering, ASCE, 111(7), pp. 1043-1056, 1985.

Wilson, E.B. and Hilferty, M.M. "The Distribution of Chi-Square", Proceedings of the National Academy of Sciences, 17, pp. 684-688, 1931.

Xu, R. and Ye, B. "Estimating Rare Floods by Geomorphological Methods", Journal of Hydrology, 96, pp. 117-124, 1987.

Yevjevich, V. "Probability and Statistics in Hydrology", Water Resources Publications, Fort Collins, Colorado, 1972.

Zheng, C.-Z. "Investigation and Regionalization of Historical Floods in China", Journal of Hydrology, 96, pp. 41-52, 1987.

APPENDIX A

COMPUTER PROGRAM VARLP3

A.1 General

This program was designed to run on DEC PRO 350 and 380 microcomputers with PRO/Tool kit FORTRAN-77.

The program queries the user for the parameters of the log Pearson Type III distribution and the sample size. If historical information is available, the program requests the number of fully defined floods greater than or equal to the censoring threshold, the number of fully defined floods below the threshold, and the number of censored floods below the threshold. The program also asks the user to input the standard normal deviate corresponding to the censoring threshold.

The program outputs the elements of the inverse-dispersion matrix and its inverse elements. The 100-year flood and its standard error of estimate in percent of the 100-year flood are also given.

A.2 List of Variables

<u>Input Variables</u>

A - scale parameter of the LP3 distribution

B - shape parameter of the LP3 distribution

M - location parameter of the LP3 distribution

T - standard normal deviate

IHIST - 0 implies conventional sample

1 implies historical information available

NB - the number of fully defined floods below the censoring threshold

NC - the number of censored floods below the threshold

N - the number of fully defined floods.

Output Variables

EYM2 - $E(y^{-2})$ as per equation (3.50)

EYM1 - $E(y^{-1})$ as per equaiton (3.49)

V(3,3) - elements of the inverse-dispersion matrix

EY - E(y) as per equation (3.24)

D - determinant of the matrix V

VARA - var(a)

VARB - var(b)

VARM - var(m)

COVAB - cov(ab)

COVAM - cov(am)

COVBM - cov(bm)

XT - the 100-year flood

SEEXT - the standard error of the 100-year flood in % of the 100-year flood.

```
      PROGRAM VARLP3
C-----------------------------------------------------------------------
C
C  VARLP3 : PROGRAM TO COMPUTE THE VARIANCE OF THE QT YEAR EVENT
C           OF A SAMPLE FROM A TRUNCATED LP3 DISTRIBUTION.
C
C  WRITTEN BY PAUL J. PILON, UNIVERSITY OF OTTAWA.
C  MAY 21, 1986.
C
C  ALL RIGHTS RESERVED.
C
C-----------------------------------------------------------------------
      REAL*8 V(3,3),D,VARA,VARB,VARM,COVAB,COVAM,COVBM,ZT,DZDA,Z,TC,
     1       DZDB,VARZT,VARXT,SEEXT,Q,NB,NC,K,XC,YC,A,B,M,XM,SD,
     2       CS,T,DM2,I10,FY,DFY,EYM2,I10M2,FYM2,I10M1,FYM1,EYM1,
     3       X10,X20,I10P1,FYP1,EY,M,X30,D2TC,DTDB,DI10DB,FFTC,U,XT
C  ENTER THE TRANSFORMED MEAN,STANDARD DEVIATION, AND SKEWNESS OF THE
C  GENERATED SAMPLE SO THAT A,B, AND M MAY BE COMPUTED FROM THE MOMENT
C  RELATIONSHIPS.
      WRITE(5,500)
      ACCEPT *,A
      WRITE(5,501)
      ACCEPT *,B
      WRITE(5,502)
      ACCEPT *,M
C  COMPUTE THE STATISTICS.
      CS=2./SQRT(B)
      IF(A.LT.0.) CS=-CS
      SD=2.*A/CS
      XM=M+2.*SD/CS
10    CONTINUE
C  COMPUTE THE CENSORING THRESHOLD XC = Q10:T=1.282.
      WRITE(5,508)
      ACCEPT *,T
      IF(A.LT.0.) T=-T
      XC=EXP(M+A*(T/(3.*B**(1./6.))-1./(9.*B**(2./3.))+
     1   B**(1./3.))**3.)
      YC=(DLOG(XC)-M)/A
C
C-----------------------------------------------------------------------
1     WRITE(5,506)
      ACCEPT *,IHIST
C  IF IHIST=0 THEN THERE IS NO HISTORIC INFORMATION.
      IF(IHIST.GT.0) THEN
      WRITE(5,503)
      ACCEPT *,NB
      WRITE(5,504)
      ACCEPT *,NC
      ENDIF
      WRITE(5,505)
      ACCEPT *,N
```

```
C----------------------------------------------------------------------
      IF (IHIST.GT.0) THEN
      K=NC
C  SOLUTION OF    2
C                d lnL
C                -----
C                  2
C                dM
C
      CALL SLPEB2(YC,B,I10,EY,A)
      Q=NC/(NB+NC)
      Z=B-2.
      CALL SLPEB2(YC,Z,I10M2,EYM2,A)
      EYM2=((1.-Q)*I10M2/((B-1.)*(B-2.))+(1.-I10M2)/
     1     ((B-1.)*(B-2.)))/(1.-Q*I10)
      EYM2=N*EYM2
      WRITE(6,700)EYM2
700   FORMAT(5X,'EYM2=',G14.7,/)
      DEY=EY*(B-1.)/YC - EY
      V(3,3)=(K/(A*A))*(DEY/I10-(EY/I10)**2)
      WRITE(6,801)V(3,3)
801   FORMAT('  CENSORED TERM OF dmdm=',G13.6,/)
      V(3,3)=(K/(A*A))*(DEY/I10-(EY/I10)**2)-(B-1.)*EYM2/(A*A)
      V(3,3)=-V(3,3)
701   FORMAT(5X,'r 3,3=',G14.7,/)
C  SOLUTION OF    2
C                d lnL
C               -------
C                dB dM
C
      Z=B-1.
      CALL SLPEB2(YC,Z,I10M1,EYM1,A)
      EYM1=((1.-Q*I10M1)/(B-1.))/(1.-Q*I10)
      EYM1=N*EYM1
      WRITE(6,702)EYM1
702   FORMAT(5X,'EYM1=',G14.7,/)
      CALL DIGAMM(B,X10)
C     X20= d lnF(Yc)
C          ---------
C          dB
C
      CALL SLOPEB(YC,B,X20,A,FETC)
      V(2,3)=-K*((EY/I10)*
     1       (DLOG(YC)-X10-X20))/A
      WRITE(6,802)V(2,3)
802   FORMAT(' CENSORED TERM OF dbdm=',G13.6,/)
C      IF(A.GT.0.) THEN
       V(2,3)=-EYM1/A-K*((EY/I10)*
     1       (DLOG(YC)-X10-X20))/A
C      ELSE
C        V(2,3)=-EYM1/A+K*((EY/I10)*
C    1         (DLOG(YC)-X10-X20))/A
```

```
C      ENDIF
      V(2,3)=-V(2,3)
703   FORMAT(5X,'r 2,3=',G14.7,/)
C  SOLUTION OF     2
C                 d  lnL
C                 ------
C                   2
C                 dA
C
      Z=B+1.
      CALL SLPEB2(YC,Z.I10P1,EYP1,A)
      EY=B*(1.-Q*I10P1)/(1.-Q*I10)
      EY=N*EY
      WRITE(6,704)EY
704   FORMAT(5X,'EY  =',G14.7,/)
      V(1,1)=-K*(YC*EY/(A*I10))*(-2./A
     1        -YC*DEY/(A*EY)+YC*EY/(A*I10))
      WRITE(6,803)V(1,1)
803   FORMAT('  CENSORED TERM OF dada='.G13.6,/)
      V(1,1)=-2.*EY/(A*A)+N*B/(A*A)-K*(YC*EY/(A*I10))*(-2./A
     1        -YC*DEY/(A*EY)+YC*EY/(A*I10))
      V(1,1)=-V(1,1)
705   FORMAT(5X,'r 1,1=',G14.7,/)
C  SOLUTION OF    2
C                d lnL
C                -----
C                dA db
C
      DTDB=(YC**(1./3.))/(2.*(B**(5./6.)))
     1     -1./(6.*B**1.5)-1.5/(B**.5)
      IF(A.LT.0.)DTDB=-DTDB
      V(1,2)=-K*(YC*EY/(A*I10))*(DLOG(YC)-X10-(EETC/I10)*DTDB)
      WRITE(6,804)V(1,2)
804   FORMAT('  CENSORED TERM OF dadb=',G13.6,/)
C      IF(A.GT.0.) THEN
        V(1,2)=-N/A-K*(YC*EY/(A*I10))*(DLOG(YC)-X10-(EETC/I10)*DTDB)
C      ELSE
C        V(1,2)=-N/A+K*(YC*EY/(A*I10))*(DLOG(YC)-X10-(EETC/I10)*DTDB)
C      ENDIF
      WRITE(6,810)I10
810   FORMAT(5X,' E(Yc) =',G14.7,/)
      V(1,2)=-V(1,2)
706   FORMAT(5X,'r 1,2=',G14.7,/)
C  SOLUTION OF    2
C                d lnL
C                -----
C                  2
C                dB
C
      CALL TRIGAM(B,X30)
C
C
```

```
      D2TC=(-5.*YC**(1./3.))/(12.*B**(11./6.))+.25/(B**2.5)+
     1      .75/(B**1.5)
      IF(A.LT.0.) D2TC=-D2TC
      TC=(YC**(1./3.)+1./(9.*B**(2./3.))-B**(1./3.))*3*B**(1./6.)
      IF(A.LT.0.) TC=-TC
      V1=1./I10
      DUDB=-TC*FFTC*DTDB
      DVDB=-FFTC*DTDB/(I10*I10)
      V(2,2)=K*(FFTC*V1*D2TC+FFTC*DTDB*DVDB+V1*DTDB*DUDB)
      WRITE(6,805)V(2,2)
805   FORMAT('  CENSORED TERM OF dbdb=',G13.6,/)
      V(2,2)=-N*X30+V(2,2)
      WRITE(6,811)D2TC,DUDB,DTDB,TC,FFTC,I10,DVDB
811   FORMAT(5X,'D2TC=',G14.7,/,5X,'DUDB=',G14.7,/,5X,'DTDB=',
     1G14.7,/,5X,'TC=',G14.7,/,5X,'FFTC=',G14.7,/,5X,'I10=',
     2G14.7,/,5X,'DVDB=',G14.7,/)
      V(2,2)=-V(2,2)
707   FORMAT(5X,'r 2,2=',G14.7,/)
C  SOLUTION OF   2
C               d lnL
C               -----
C               dM dA
C
      V(1,3)=K*(YC*DEY/(A*A*I10)-((EY/I10)**2)*
     1        YC/(A*A)+EY/(A*A*I10))
      WRITE(6,806)V(1,3)
806   FORMAT('  CENSORED TERM OF dadm=',G13.6,/)
      V(1,3)=-N/(A*A)+K*(YC*DEY/(A*A*I10)-((EY/I10)**2)*
     1        YC/(A*A)+EY/(A*A*I10))
      V(1,3)=-V(1,3)
708   FORMAT(5X,'r 1,3=',G14.7,/)
C
      ELSE
        CALL TRIGAM(B,X30)
        V(1,1)=N*B/(A*A)
        V(1,2)=N/A
        V(1,3)=N/(A*A)
        V(2,2)=N*X30
        V(2,3)=N/(A*(B-1.))
        V(3,3)=N/(A*A*(B-2.))
      ENDIF
      WRITE(6,705)V(1,1)
      WRITE(6,706)V(1,2)
      WRITE(6,708)V(1,3)
      WRITE(6,707)V(2,2)
      WRITE(6,703)V(2,3)
      WRITE(6,701)V(3,3)
C  NOW FILL IN THE REST OF THE MATRIX
      V(3,1)=V(1,3)
      V(2,1)=V(1,2)
      V(3,2)=V(2,3)
C
```

```
C  THE INVERSE DISPERSION MATRIX IS NOW COMPLETE.
C  COMPUTE THE DETERMINANT OF THE MATRIX.
      D=V(1,1)*V(2,2)*V(3,3)-V(1,1)*V(2,3)*V(3,2)-V(1,2)*V(2,1)*
     1  V(3,3)+V(1,2)*V(2,3)*V(3,1)+V(1,3)*V(2,1)*V(3,2)-
     2  V(1,3)*V(2,2)*V(3,1)
      WRITE(6,709)D
709   FORMAT(5X,'DET.=',G14.7,/)
C
      VARA=(V(2,2)*V(3,3)-V(3,2)**2)/D
      VARB=(V(1,1)*V(3,3)-V(1,3)**2)/D
      VARM=(V(1,1)*V(2,2)-V(1,2)**2)/D
      WRITE(6,710)VARA,VARB,VARM
710   FORMAT(5X,'VARA=',G14.7,/,5X,'VARB=',G14.7,/,
     1       5X,'VARM=',G14.7,/)
C
      COVAB=-1*(V(1,2)*V(3,3)-V(1,3)*V(3,2))/D
      COVAM=(V(2,1)*V(3,2)-V(3,1)*V(2,2))/D
      COVBM=-1*(V(1,1)*V(3,2)-V(3,1)*V(1,2))/D
      WRITE(6,711)COVAB,COVAM,COVBM
711   FORMAT(5X,'COVAB=',G14.7,/,5X,'COVAM=',G14.7,/,5X,
     1       'COVBM=',G14.7,/)
C  ln XT=ZT; DZ/DA
      T100=2.326
      IF (A.LT.0.0) T100=-T100
      U=(T100/(3.*B**(1./6.))-1./(9.*B**(2./3.))+B**(1./3.))
      DZDA=U**3.
C  DZ/DB
      DZDB=(3.*A*U**2.)*(-T100/(18.*B**(7./6.))+2./(27.*
     1     B**(5./3.))+1./(3.*B**(2./3.)))
      VARZT=DZDA*DZDA*VARA+DZDB*DZDB*VARB+VARM+2.*DZDA*DZDB*COVAB
     1     +2.*DZDA*COVAM+2.*DZDB*COVBM
      ZT=M+A*DZDA
      XT=EXP(ZT)
      SEEXT=(SQRT(VARZT))*100.
C  OUTPUT S.E.E. AS EVERYTHING HAS BEEN COMPUTED!
C
      WRITE(6,600)XT,SEEXT,A,B,M
      WRITE(5,507)
      ACCEPT *,IC
      IF(IC.GT.0) GO TO 1
      STOP
C  FORMAT STATEMENTS
500   FORMAT(3X,' ENTER A, THE SCALE PARAMETER=?',/)
501   FORMAT(3X,' ENTER B, THE SHAPE PARAMETER=?',/)
502   FORMAT(3X,' ENTER M, THE LOCATION PARAMETER=?',/)
503   FORMAT(3X,' # OF OBS. FLOODS < THRESHOLD, NB=?',/)
504   FORMAT(3X,' # OF CENSORED FLOODS        , NC=?',/)
505   FORMAT(3X,' # OF OBS. FLOODS             , N=?',/)
506   FORMAT(3X,' HISTORICAL ANALYSIS?',/,3X,' 1 FOR YES',/,
     13X,' 0 FOR NO.....',/)
507   FORMAT(3X,' START OVER?',/,3X,' 1 FOR YES',/,3X,
     1'0 FOR NO.....',/)
```

```
508   FORMAT(3X,' ENTER Z, 1/10=1.282, 1/166.7=2.51, 1/63=2.145:'
     1,/)
600   FORMAT(' Q100=',G14.7,3X,' S.E.E. Q100=',G14.7,/,
     1' PARAMETERS ARE:  A=',G14.7,3X,' B=',G14.7,3X,' M=',G14.7,/)
      END
      SUBROUTINE TRIGAM(B,X30)
C
C     X30 IS THE TRIGAMMA OF B
      REAL*8 B,X30
      X30=1./(B+2)+1./(2.*(B+2)**2)+1./(6.*(B+2)**3)
     1    -1./(30.*(B+2)**5)+1./(42.*(B+2)**7)-1./(30.*(B+2)**9)
     2    +1./(B+1)**2+1/B**2
      RETURN
      END
      SUBROUTINE SLPEB2(YC,B,I10,FFY,A)
      REAL*8 YC,B,I10,FFY,A,TC,AX,FTC,XLNGB
C  YC AND B ARE TRANSFERRED FROM THE CALLING PROGRAM.
C
C  I10=F(YC)
C  FFY=F(YC)
C
      TC=3.*YC**(1./3.)*B**(1./6.)+1./(3.*B**.5)-3.*B**.5
      IF (ABS(TC).GT.20.) THEN
         FFY=0.
      I10=1.
      GO TO 100
       ENDIF
      AX=(0.17401/(1+0.33267*ABS(TC))-0.04794/(1+0.33267*ABS(TC))**2
     & +0.37393/(1+0.33267*ABS(TC))**3)*EXP(-TC**2/2)
      FTC=1.-AX
      IF (A.GT.0..AND.TC.LT.0.) FTC=AX
      IF (A.LT.0..AND.TC.GT.0.) FTC=AX
      I10=FTC
      TC=2.*3.14159265
      XLNGB=DLOG(1.+1./(12.*B))+(B-.5)*DLOG(B)+.5*DLOG(TC)-B
      FFY=EXP(0.-YC+(B-1.)*DLOG(YC)-XLNGB)
      IF (A.LT.0.) FFY=-FFY
100   RETURN
      END
      SUBROUTINE DIGAMM(B,X10)
C  SUBROUTINE TO COMPUTE THE DIGAMMA FUNCTION OF B.
C  X10= DIGAMMA(B)
      REAL*8 B,X10
      X10=DLOG(B+2.)-1./(2.*(B+2.))-1./(12.*(B+2.)**2)+
     1    1./(120.*(B+2.)**4)-1./(252.*(B+2.)**6.)-1./(B+1.)-1./B
      RETURN
      END
C
      SUBROUTINE SLOPEB(YC,B,X20,A,FFTC)
C  YC AND B ARE TRANSFERED FROM THE CALLING PROGRAM.
C  X20 = D/DB OF LNF(YC)
      REAL*8 YC,B,X20,A,FFTC,TC,AX,FTC
```

```
      TC=3.*YC**(1./3.)*B**(1./6.)+1./(3.*B**.5)-3.*B**.5
      IF (ABS(TC).GT.20) THEN
         X20=0.
         GO TO 100
      ENDIF
      FFTC=1/(2*3.14159265)**.5*EXP(-TC**2/2.)
C      IF (A.LT.0.) FFTC=-FFTC
      AX=(0.17401/(1+0.33267*ABS(TC))-0.04794/(1+0.33267*ABS(TC))**2
     & +0.37393/(1+0.33267*ABS(TC))**3)*EXP(-TC**2/2)
      FTC=1.-AX
      IF(A.GT.0..AND.TC.LT.0.) FTC=AX
      IF(A.LT.0..AND.TC.GT.0.) FTC=AX
      X20=YC**(1./3.)/(2.*B**(5./6.))-1./(6.*B**1.5)-3./(2.*B**.5)
      X20=X20*FFTC/FTC
      IF(A.LT.0.) X20=-X20
100   RETURN
      END
C
```

APPENDIX B

COMPUTER PROGRAM LP3NET

B.1 General

This program was designed to run on a CDC CYBER 730 computer with FORTRAN V.

Program LP3NET provides a hydrologic flood frequency analysis of a typical sample, a sample with zeros, a sample with low outliers, a sample with historic information, or any combination of the latter three cases. The LP3 distribution is used where the parameters are estimated by maximum likelihood. For samples with zeros and/or low outliers, the method is modified by using synthetic statistics (Hydrology Subcommittee 1982; Pilon et al. 1985); for samples with historic information, the sample is treated as a censored sample from an LP3 distribution. Flows corresponding to selected return periods are computed from the probability function.

The program requires input the sample series of hydrologic events and their date of occurrence, including those which have been identified as low outliers, and any historical information.

The output consists of input data, ranked data and adjusted ranking if historic information is available, low outliers and zeros identified if present, empirical probabilities, and return periods. Then follow the sample's statistics, the distribution's parameters, and tabular frequency regime, with an optional plot of the cumulative density function and the data.

B.2 Input Format for Peak Flow Analysis

<u>Run Control Card</u>

Prepare run control cards as follows:

Cols 1-7 Format A7- Water Survey of Canada station number for which analysis is required

Cols 8-9 Blank

Col 10 Format I1 - Plotting requirement:

0 - No plot required

1 - Plot required

Cols 11-12 Blank

Col 13 Format I1 - Number of nonzero low outliers identified

Cols 14-15 Blank

Col 16 Format I1 - Type of analysis required:

0 - Conventional LP3

1 - Historical LP3

Cols 17-18 Blank

Col 19 Format I1 - Number of zero flows in record

Cols 20-21 Blank

Col 22 Format I1 - Number of high outliers identified

Repeat the above for as many stations as desired.

Data required for each run

(1) First parameter card

Cols 1-7 Format A7 - Water Survey of Canada station number

(2) Second parameter card

Cols 1-59 Format 5A10 & A9 - Water Survey of Canada station number and name (for title)

Col 60 Format I1 - Metric or Imperial measure specification:

1 - Imperial units (ft^3/s)

2 - Metric units (m^3/s)

Cols 77-78 Format I2 - Sample size

(3) Peak flod data cards

Cols 1-10 Format A10 - Water Survey of Canada station number (optional)

Cols 11-13 Format I3 - Year of data sample less 1000
(e.g. 1965: code as 965)

Cols 14-15 Format I2 - Month in which peak flow occurred

Cols 16-21 Format F6.0 - Peak flow (up to 5 figures after decimal)

Cols 22-32
Cols 11-21
Cols 33-43
Cols 44-54
Cols 55-65
Cols 66-76

Repeat Cols. 11-21 with subsequent years, months and peak flow data. Use as many cards as necessary to include all data. Note that there are 6 sets of data per card.

(4) Historical Analysis Parameter Card

This card is required only when performing an historical LP3 analysis.

Cols 3-5 Format I3 - Total time span, YT

Cols 11-20 Format F10.0 - Censoring threshold

Cols 24-25 Format I2 - Number of peak flow data above or equal to the threshold value.

Repeat steps (1) to (4) for each station to be run.

```
      PROGRAM LP3NET(INPUT,OUTPUT,TAPE30,TAPE50,TAPE6=OUTPUT,
     1 TAPE5=INPUT)
C
C  THIS PROGRAM PERFORMS A FLOW FREQUENCY ANALYSIS USING THE LOG
C  PEARSON TYPE III DISTRIBUTION.  LP3NET IS VERSATILE ENOUGH TO
C  ANALYZE EXTREME SERIES HAVING NO OUTLIERS, OR THOSE HAVING LOW AND/OR
C  HIGH OUTLIERS, AND THOSE HAVING ZERO FLOWS. THE ASYMPTOTIC STANDARD
C  ERROR OF ESTIMATE AND THE K-FACTOR ARE COMPUTED FOR THE T-YEAR EVENT.
C  THE PROGRAM ALSO HAS THE ABILITY TO FIT THE LP3 DISTRIBUTION TO
C  LOW FLOWS.
C
C  WRITTEN BY PAUL J. PILON
C  UNIVERSITY OF OTTAWA
C  OTTAWA, ONTARIO.
C
C  AUGUST 1987
C
C  *** ALL RIGHTS RESERVED ****
C  THIS PROGRAM IS COPYRIGHTED AND NO PORTION SHALL BE COPIED WITHOUT
C  THE WRITTEN PERMISSION OF THE AUTHOR.
      CHARACTER*11 EXM1
      INTEGER HIST,YT.FAIL,HOUT
      REAL M,LNQ(150),LNMEAN,LNSD,LNCV,LNG1.LNQBAR,LNSTDV,LNCOV,LNSKEW
      REAL MM,LNCK,QD(3)
      DIMENSION RP(11),FLOOD(11),BUFF(1024),PERIOD(3)
      DIMENSION Q(3),VMAT(3,3)
      COMMON/BLOCK1/FLOW(150),PROB(150),STN(6),MN,IYR(150),IMON(150),LOU
     1T,DFLOW(150),DFLOW1(150),HOUT,INDIC,IPRD(2),MF
      COMMON/BLOCK2/QBAR,N,NET,KOUT
      COMMON/BLOCK4/A,B,M
      COMMON/BLOCK6/NHA,NB,NC,THRESH,YT,UNIT(2).XNA
C
      DATA(RP(I),I=1,11)/1.003,1.05,1.25,2.,5.,10.,20.,50..100.,200.,500
     1./
      DATA(PERIOD(I),I=1,3)/2.,10.,100./
      DATA(UNIT(I),I=1,2)/3HCFS,3HCMS/
C
C  INPUT FORMAT STATEMENTS...
  5   FORMAT(A7,6(1X,I2))
  6   FORMAT(5A10,A9,I1,16X,I2)
  7   FORMAT(10X,6(I3,I2,F6.0))
  8   FORMAT(I5,5X,F10.0,I5)
  9   FORMAT(5A10,A9,I1,A10,A6,I2,I3)
C
C  OUTPUT FORMAT STATEMENTS...
 10   FORMAT('0',2X,'RUN ABORTED - FLOW RECORDS FOR',1X,A7,1X,'ARE MISSI
     1NG FROM THE MASTER FILE',/,'0')
 11   FORMAT('1',///, 8X,'FREQUENCY ANALYSIS - LOG PEARSON TYPE III ',
     1'DISTRIBUTION',//,12X,6A10)
 12   FORMAT('1',///, 3X,'HISTORICAL FREQUENCY ANALYSIS - LOG PEARSON '.
     1'TYPE III DISTRIBUTION',//,12X,6A10)
```

```
13   FORMAT(///,32X,'SAMPLE STATISTICS',//,
    1 23X,'MEAN',7X,'S.D.',8X,'C.V.',7X,'C.S.',7X,'C.K.',/,
    2 6X,'    X SERIES ',F10.3,2X,F9.3,4X,F7.3,5X,F6.3,5X,F6.3)
14   FORMAT(//,19X,'AFTER REMOVAL OF ZEROES AND/OR LOW OUTLIERS',//,
    1 23X,'MEAN',7X,'S.D.',8X,'C.V.',7X,'C.S.',7X,'C.K.',/,
    2 6X,'    X SERIES ',F10.3,2X,F9.3,4X,F7.3,5X,F6.3,5X,F6.3,/,
    3 6X,'LN X SERIES ',F10.3,2X,F9.3,4X,F7.3,5X,F6.3,5X,F6.3)
15   FORMAT(//,21X,'SOLUTION OBTAINED VIA MAXIMUM LIKELIHOOD',//)
16   FORMAT(//,21X,'SOLUTION OBTAINED VIA MOMENTS',//)
17   FORMAT(7X,'MOMENTS FROM PARAMETERS: MEAN=',F8.3,2X,'S.D.=',
    1 F7.3,2X,'C.S.=',F6.3,/)
18   FORMAT(7X,'LP3 PARAMETERS:',1X,'A=',G11.4,1X,'B=',G11.4,1X,
    &'LOG(M)=',G11.4,/,55X,'M =',A11,/)
19   FORMAT(8X,'PARAMETERS OF THE LP3 WHICH DUPLICATES THE CONDITIONAL
    & FUNCTION:',/,9X,'A=',G11.4,1X,'B=',G11.4,1X,'LOG(M)=',G11.4,
    & 1X,'M =',A11,/)
20   FORMAT(9X,'SYNTHETC STATISTICS: MEAN=',F8.3,2X,'S.D.=',F7.3,1X,
    & 'C.S.=',F7.3,/)
21   FORMAT(6X,'LN X SERIES ',F10.3,2X,F9.3,4X,F7.3,5X,F6.3,5X,F6.3)
22   FORMAT(/,6X,'X(MIN)= ',F9.3,1X,A3,23X,'TOTAL SAMPLE SIZE= ',I3,
    1/,6X,'X(MAX)= ',F9.3,1X,A3,21X,'NO. OF LOW OUTLIERS= ',
    2I3,/,50X,'NO. OF ZERO FLOWS= ',I3,/)
23   FORMAT(15X,'ZERO FLOW IN SERIES - LN X SERIES CANNOT BE COMPUTED')
24   FORMAT(/,6X,'X(MIN)= ',F9.3,1X,A3,23X,'TOTAL SAMPLE SIZE= ',I3,
    1/,6X,'X(MAX)= ',F9.3,1X,A3,21X,'NO. OF LOW OUTLIERS= ',
    2I3,/,50X,'NO. OF ZERO FLOWS= ',I3,/,47X,'NO. OF DELETED FLOWS= ',
    3I3,/)
25   FORMAT(9X,F7.3,10X,F5.3,14X,1H-,9X,1H-,11X,1H-)
29   FORMAT(///,30X,'FLOW FREQUENCY REGIME',//,10X,
    1'RETURN',8X,'EXCEEDANCE',9X,'FLOW',5X,'STANDARD',3X,'K-FACTOR',
    2/,10X,'PERIOD',8X,'PROBABILITY',8X,'(',A3,')',4X,'ERROR(%)',
    33X,'FOR GLS',/)
30   FORMAT(6X,'NOTE: ',I3,' ITERATIONS REQUIRED TO OBTAIN A ',
    & 'POSITIVE DETERMINANT.',/)
32   FORMAT(16X,'DISTRIBUTION IS UPPER BOUNDED AT M =',A11)
33   FORMAT(16X,'NOTE: THE MAXIMUM OBSERVED FLOW EXCEEDS THE',
    1' UPPER',/,22X,'BOUNDARY OF THE DISTRIBUTION. THUS, THE',
    2' DESIGN ',/,22X,'DISCHARGES ARE NOT PRINTED.',//)
34   FORMAT('0',2X,'RUN ABORTED - TAPE 30 INFORMATION FOR',1X,
    1A7,1X,'IS INCORRECT.',/,3X,'LOW FLOW HISTORICAL ANALYSIS',
    2' IS NOT AVAILABLE.',/,'0')
35   FORMAT(//,30X,'FLOW FREQUENCY REGIME',//,27X,'NON',/,10X,
    1'RETURN',8X,'EXCEEDANCE',9X,'FLOW',5X,'STANDARD',3X,'K-FACTOR',
    2/,10X,'PERIOD',8X,'PROBABILITY',8X,'(',A3,')',4X,'ERROR(%)',
    33X,'FOR GLS',/)
40   FORMAT(9X,F7.3,10X,F5.3,8X,F10.0,4X,F6.2,5X,F7.4)
41   FORMAT(9X,F7.3,10X,F5.3,8X,F10.1,4X,F6.2,5X,F7.4)
42   FORMAT(9X,F7.3,10X,F5.3,8X,F10.2,4X,F6.2,5X,F7.4)
43   FORMAT(9X,F7.3,10X,F5.3,8X,F10.3,4X,F6.2,5X,F7.4)
44   FORMAT(9X,F7.3,10X,F5.3,8X,F10.0,6X,1H-,8X,F7.4)
45   FORMAT(9X,F7.3,10X,F5.3,8X,F10.1,6X,1H-,8X,F7.4)
46   FORMAT(9X,F7.3,10X,F5.3,8X,F10.2,6X,1H-,8X,F7.4)
```

```
 47    FORMAT(9X,E7.3,10X,F5.3,8X,F10.3,6X,1H-,8X,E7.4)
       NPLT=0
C  READ THE CONTROL CARDS FROM THE TWO INPUT FILES AND CHECK THAT THE
C  DATA ARE AVAILABLE...
100    READ(30,5,END=999) ISTA,IPLOT,LOWS,HIST,NZERO,HOUT,INDIC
       REWIND 50
102    READ(50,5,END=103) NSTA
       IF(NSTA.NE.ISTA) GO TO 102
       IF(INDIC.EQ.1.AND.HIST.EQ.1) THEN
         WRITE(6,34) ISTA
         GOTO 100
       ENDIF
       IF(IPLOT.NE.0) NPLT=NPLT+1
       IF(IPLOT.NE.0.AND.NPLT.EQ.1) CALL PLOTS(BUFF,1024)
       GO TO 105
103    WRITE(6,10) ISTA
       GO TO 100
105    CONTINUE
C
       LOUT=LOWS+NZERO
C
C  READ THE STATION NAME AND PEAK FLOW RECORDS...
       IF(INDIC.EQ.1) THEN
         READ(50,9) (STN(I),I=1,6),MN,(IPRD(J),J=1,2),N,MF
       ELSE
         READ(50,6) (STN(I),I=1,6),MN,N
       ENDIF
       DO 110 II=1,N,6
       IL=II+5
       READ(50,7) (IYR(I),IMON(I),FLOW(I),I=II,IL)
110    CONTINUE
       N1=N
       IF(HIST.EQ.0) GO TO 120
C  READ IN ADDITIONAL INFO FOR HISTORICAL FLOOD ANALYSIS...
       READ(50,8) YT,THRESH,NHA
C  SORT AND WRITE FLOWS FOR HISTORICAL ANALYSIS...
       CALL LISTS
       GO TO 130
C
120    CONTINUE
C  SORT AND WRITE FLOWS IF NO HISTORICAL ANALYSIS IS REQUIRED...
       CALL WRITES
130    CONTINUE
C
C  COMPUTE AND WRITE NATURAL LOGS AND STATISTICS...
C   IF ZERO FLOWS ARE PRESENT, LOGS ARE NOT COMPUTED
       CALL STATS(DFLOW,N1,QMEAN,SD,CV,G1,CK)
       DO 140 I=1,N
       IF(DFLOW(I).EQ.0.) GO TO 145
       LNQ(I)=ALOG(DFLOW(I))
140    CONTINUE
       CALL STATS(LNQ,N1,LNMEAN,LNSD,LNCV,LNG1,LNCK)
```

```
145   CONTINUE
      IF(HIST.EQ.0) WRITE(6,11) (STN(J),J=1,6)
      IF(HIST.GT.0) WRITE(6,12) (STN(J),J=1,6)
      WRITE(6,13)QMEAN,SD,CV,G1,CK
      IF(DFLOW(N).EQ.0.) WRITE(6,23)
      IF(DFLOW(N).GT.0.) WRITE(6.21) LNMEAN.LNSD.LNCV,LNG1,LNCK
      WM=LNMEAN
      SW=LNSD
      CS=LNG1
146   IF (LOUT.GT.0) WRITE(6,22) DFLOW(N),UNIT(MN).N,
     & DFLOW(1),UNIT(MN),LOWS,NZERO
      NET=N-LOUT
      QBAR=QMEAN
      SKEW=G1
      STDEV=SD
      IF(LOUT.EQ.0) GO TO 150
C  IF THERE ARE LOW OUTLIERS, RE-COMPUTE STATS WITHOUT THEM...
      N1=NET
      CALL STATS(DFLOW,N1,QBAR,STDEV,COV.SKEW,CK)
      CALL STATS(LNQ,N1,LNQBAR,LNSTDV,LNCOV,LNSKEW,LNCK)
      WRITE(6,14) QBAR,STDEV,COV,SKEW,CK,LNQBAR,LNSTDV.
     1 LNCOV,LNSKEW,LNCK
      WM=LNQBAR
      SW=LNSTDV
      CS=LNSKEW
150   CONTINUE
C
C  COMPUTE THE THREE PARAMETERS OF THE LOG PEARSON TYPE 3 DIST.
      KOUT=0
      IFAIL=0
C  FIRST TRY THE MAXIMUM LIKELIHOOD METHOD OF SOLUTION.
C CHECK FOR HISTORIC INFORMATION. HIST=1.
      IF (HIST.EQ.0) THEN
        CALL MAXLP3 (LNQ,CS,IFAIL)
C  IF SOLN OBTAINED THEN IFAIL=0, IF NO SOLN THEN IFAIL=1.
        IF (IFAIL.GT.0) THEN
          A=(SW*CS)/2.
          B=(2./CS)**2.
          M=WM-2.*SW/CS
        ENDIF
      ENDIF
      IF (HIST.GT.0) THEN
C  ATTEMP MAXIMUM LIKELIHOOD SOLN FIRST
        CALL LP3HIS(LNQ,IFAIL)
        IF (IFAIL.GT.0) THEN
          CALL SOL3HW(LNQ,WM,SW,CS,CVM)
          KOUT=0
          A=(SW*CS)/2.
          B=(2./CS)**2
          M=WM-2.*SW/CS
        ELSE
          IF (KOUT.GT.0) THEN
```

```
            NET=NET-KOUT
            N1=NET
            CALL STATS (DFLOW,N1,QBAR,STDEV,COV.SKEW,CK)
            CALL STATS(LNQ,N1,LNQBAR,LNSTDV.LNCOV,LNSKEW,LNCK)
            WRITE(6,14) QBAR,STDEV,COV,SKEW,CK.LNQBAR,LNSTDV.
     1                  LNCOV,LNSKEW,LNCK
            NET=N-LOUT
            WRITE(6,24) DFLOW(N),UNIT(MN),N,DFLOW(1),UNIT(MN),
     1                  LOWS,NZERO,KOUT
          ENDIF
        ENDIF
      ENDIF
C  WRITE OUT THE SOLUTION METHOD.
      IF (HIST.EQ.0) THEN
        IF (IFAIL.EQ.0) THEN
          WRITE(6,15)
        ELSE
          WRITE(6,16)
        ENDIF
      ELSE
        IF(IFAIL.EQ.0) THEN
          WRITE(6,15)
        ELSE
          WRITE(6,16)
        ENDIF
      ENDIF
C  RETRO FIT A,B,M IF LOUT.GT.0
      ADJUST=FLOAT(N)/NET
      IF (LOUT.GT.0) THEN
C  RETRO-FIT OF THE LP3 AS PER 17B.
        DO 160 I=1,3
        PR=1./PERIOD(I)
        PRADJ=PR*ADJUST
        PRNEW=PRADJ
        IF(PRADJ.GT.0.5)PRNEW=1.-PRADJ
        PRXTRA=PRADJ
        RET=1./PRNEW
        T=SQRT(ALOG(RET*RET))
        X1=T-(2.51552+0.80285*T+0.01033*T*T)/
     &     (1.+1.43279*T+0.18927*T*T+0.00131*T*T*T)
        IF(PRXTRA.GT.0.5)X1=0.-X1
        IF(A.LT.0.)X1=-X1
        X4=M+A*(X1/(3.*(B**(1./6)))-1./(9.*(B**(2./3)))
     &     +B**(1./3))**3
        QD(I)=EXP(X4)
 160    CONTINUE
C  COMPUTE SYNTHETIC STATISTICS
        XCS=-2.5+3.12*(ALOG(QD(3)/QD(2))/ALOG(QD(2)/QD(1)))
        XK01=(2./XCS)*(((((2.326-XCS/6.)*XCS/6.)+1.)**3)-1.)
        XK5=(2./XCS)*(((((0.-XCS/6.)*XCS/6.)+1.)**3.)-1.)
        SW=ALOG(QD(3)/QD(1))/(XK01-XK5)
        WM=ALOG(QD(1))-XK5*SW
```

```
C  COMPUTE NEW A,B,M
         A=(SW*XCS)/2.
         B=(2./XCS)**2.
         M=WM-2.*SW/XCS
       ENDIF
C  RETRO OR NO RETRO, COMPUTE M IN REAL SPACE.
 445   IF(ABS(M).LE.88.) THEN
         EXM=EXP(M)
         ENCODE(11,446,EXM2)EXM
         DECODE(11,447,EXM2)EXM1
 446     FORMAT(G11.4)
 447     FORMAT(A11)
       ELSE
         IF(M.GT.88.)EXM1='+ INFINITY '
         IF(M.LT.-88.)EXM1=' ZERO      '
       ENDIF
C  PARAMETERS HAVE BEEN RECALCULATED IF LOUT >0; IF NOT THEN
C  THE OLD PARAMETER VALUES REMAIN.
C  NOW WRITE THE PARAMTERS.
       IF (A.LT.0.) WRITE (6,32)EXM1
       IF (LOUT.GT.0) THEN
         WRITE(6,19) A,B,M,EXM1
         WRITE(6,20) WM,SW,XCS
       ELSE
         WRITE(6,18) A,B,M,EXM1
C  AS LOUT=0 THERE IS NO RETRO FIT, THUS
C  IF(IFAIL.EQ.0.OR.HIST.GT.0) THEN RECOMPUTE THE MOMENTS
C  HOWEVER, IF(FAIL.GT.0.AND.HIST.EQ.0) THEN THERE IS NO
C  NEED TO RECOMPUTE THE STATS AS A,B,M ARE DERIVED FROM THEM.
         IF (IFAIL.EQ.0.OR.HIST.GT.0) THEN
           CS1=2./SQRT(B)
           IF(A.LT.0.) CS1=-CS1
           SW1=A*2./CS1
           WM1=M+2.*SW1/CS1
           WRITE(6,17) WM1,SW1,CS1
         ENDIF
       ENDIF
       IF (LNQ(1).GT.M.AND.A.LT.0.) THEN
         WRITE(6,33)
         GOTO 998
       ENDIF
C  COMPUTE THE VARIANCES AND COVARIANCES OF THE PARAMETERS
C  USING THE M.L. APPROACH. MOMENTS ARE USED LATER AS A BACKUP.
       IT=0
       IF(B.GT.2.) THEN
         IF(HIST.EQ.0.AND.IFAIL.EQ.0) CALL VARMLC(VARA,VARB,VARM,
      1                               COVAB,COVAM,COVBM)
         IF(HIST.EQ.1) THEN
           CALL VARMLH(VARA,VARB,VARM,COVAB,COVAM,COVBM,IT)
           IF(IT.GT.0) WRITE(6,30)IT
         ENDIF
       ENDIF
```

```
      IF (INDIC.EQ.0) THEN
        WRITE(6,29) UNIT(MN)
      ELSE
        WRITE(6,35) UNIT(MN)
      ENDIF
C  COMPUTATIONS OF DESIGN FLOODS
      DO 460 I=1,11
      PR=1./RP(I)
      PRNEW=PR
      IF(LOUT.GT.0.AND.(1./ADJUST).LE.PR) THEN
        PEAK=-999.99
        GOTO 455
      ENDIF
      IF(PR.GT.0.5)PRNEW=1.-PR
      PRXTRA=PR
      RET=1./PRNEW
      T=SQRT(ALOG(RET*RET))
      X1=T-(2.51552+0.80285*T+0.01033*T*T)/
     &   (1.+1.43279*T+0.18927*T*T+0.00131*T*T*T)
      IF(PRXTRA.GT.0.5)X1=0.-X1
      IF(A.LT.0.)X1=-X1
C  CHANGE SIGN OF X1 FOR LOW FLOW ANALYSIS
      IF(INDIC.EQ.1) X1=-X1
      U=(X1/(3.*(B**(1./6)))-1./(9.*(B**(2./3)))
     &   +B**(1./3))
      X4=M+A*(U**3)
      FLO=EXP(X4)
      CALL ROUNDS(FLO,PEAK)
      CS1=2./SQRT(B)
      IF (A.LT.0.) CS1=-CS1
      SW1=A*2./CS1
      WM1=M+2.*SW1/CS1
C       IF ((A.GT.0..AND.X1.LT.0.).OR.(A.LT.0..AND.X1.GT.0.)) GOTO 455
      IF ((B.LE.2..OR.IT.GT.8).OR.(HIST.EQ.0.AND.IFAIL.EQ.1)) THEN
C  UES OF BOBEE'S TECHNIQUE
        DKDCS=(X1**2-1)/6.+4*(X1**3.-6.*X1)*CS1/216.-3.*
     1         (X1*X1-1)*CS1*CS1/216.+4.*X1*CS1**3/1296.-
     2         10.*(CS1**4)/(6.**6)
        XK=X1+(X1*X1-1)*CS1/6.+(X1**3-6.*X1)*((CS1/6.)**2)/3.-
     1     (X1*X1-1)*(CS1/6.)**3+X1*(CS1/6.)**4+(1./3.)*(CS1/6.)
     2     **5
        VARFLO=(SW1*SW1/NET)*(1.+XK*CS1+XK**2*(3.*CS1**2/4.+1.)/2.
     1          +3.*XK*DKDCS*(CS1+(CS1**3)/4.)+(3.*DKDCS**2)*(2.+
     2          3.*CS1**2+5.*CS1**4/8.))
      ELSE
        DZDA=U**3
        DZDB=(3.*A*U**2.)*(-X1/(18.*B**(7./6.))+2./(27.*
     1     B**(5./3.))+1./(3.*B**(2./3.)))
        VARFLO=DZDA*DZDA*VARA+DZDB*DZDB*VARB+VARM+2.*DZDA*DZDB*COVAB
     1      +2.*DZDA*COVAM+2.*DZDB*COVBM
      ENDIF
      IF(INDIC.EQ.1) X1=-X1
```

```
      IF(VARFLO.GT.0.) THEN
        STERR=100.*SQRT(VARFLO)
      ELSE
        STERR=-999.
      ENDIF
C  COMPUTE K FACTOR FOR GLSNET
      XFACT=(X4-WM1)/SW1
455   IF(PEAK.LT.0) THEN
        FLOOD(I)=-999.99
        WRITE(6,25)RP(I),PR
      ELSE
        FLOOD(I)=PEAK
        IF (STERR.GT.0.) THEN
            IF (PEAK.GE.100.) THEN
                WRITE(6,40) RP(I),PR,PEAK,STERR,XFACT
              ELSE IF (PEAK.GE.10.) THEN
                WRITE(6,41) RP(I),PR,PEAK,STERR,XFACT
              ELSE IF (PEAK.GE.1.) THEN
                WRITE(6,42) RP(I),PR,PEAK,STERR,XFACT
              ELSE
                WRITE(6,43) RP(I),PR,PEAK,STERR,XFACT
            ENDIF
          ELSE
            IF (PEAK.GE.100.) THEN
               WRITE(6,44) RP(I),PR,PEAK,XFACT
              ELSE IF (PEAK.GE.10.) THEN
               WRITE(6,45) RP(I),PR,PEAK,XFACT
              ELSE IF (PEAK.GE.1.) THEN
               WRITE(6,46) RP(I),PR,PEAK,XFACT
              ELSE
               WRITE(6,47) RP(I),PR,PEAK,XFACT
            ENDIF
        ENDIF
      ENDIF
460   CONTINUE
C
C  FOR FLOOD FREQUENCY PLOTS...
        IF(IPLOT.EQ.0) GO TO 100
        IF (INDIC.EQ.0) THEN
          CALL FPLOT(FLOOD,RP,HIST,IFAIL)
        ELSE
          CALL GPLOT(FLOOD,RP)
        ENDIF
        CALL PLOT(11.0,0.0,-3)
C
998   CONTINUE
      GO TO 100
999   CONTINUE
       IF(NPLT.NE.0) CALL PLOT(0.,0.,999)
      STOP
      END
C
```

```
      SUBROUTINE WRITES
C
C  THIS S/R PRINTS DATA WITH CUMULATIVE PROBABILITIES AND RETURN
C  PERIODS, AND NOTES ANY OUTLIERS BY AN ASTERISK.
C
      INTEGER HOUT
      COMMON/BLOCK1/FLOW(150),PROB(150),STN(6),MN,IYR(150),IMON(150),LOU
     1T,DFLOW(150),DFLOW1(150),HOUT,INDIC,IPRD(2),ME
      COMMON/BLOCK2/QBAR,N,NO,KOUT
C
      NZERO=0
      ILINE=65
      DO 10 I=1,N
      DFLOW(I)=FLOW(I)
      DFLOW1(I)=FLOW(I)
      IF(FLOW(I).EQ.0.0) NZERO=NZERO+1
 10   CONTINUE
      CALL SORTD(N,DFLOW)
      IF(INDIC.EQ.1)CALL SORTX(N,DFLOW1)
      NO=N-LOUT
C
      DO 20 M=1,N
      PROB(M)=(M-0.4)/(N+0.2) * 100.
      IF(INDIC.EQ.0) THEN
        IF(DFLOW(M).EQ.0.0) PROB(M)=(N-NZERO)/FLOAT(N)*100.
      ELSE
        IF(DFLOW1(M).EQ.0.0) PROB(M)=(N-NZERO)/FLOAT(N)*100.
      ENDIF
      TIME=100./PROB(M)
      IF(ILINE.LT.60) GO TO 15
      WRITE(6,100) (STN(J),J=1,6)
      IF(INDIC.EQ.1) THEN
        WRITE(6,201) ME,(IPRD(J),J=1,2)
        WRITE(6,205) ME
      ELSE
        WRITE(6,105)
      ENDIF
      IF(MN.EQ.1) WRITE(6,110)
      IF(MN.EQ.2) WRITE(6,115)
      ILINE=12
 15   CONTINUE
      ISTAR=1H
      IF(INDIC.EQ.0) THEN
        IF(M.GT.NO.OR.M.LE.HOUT) ISTAR=1H*
      ELSE
        IF(M.LE.LOUT.OR.M.GT.(N-HOUT)) ISTAR=1H*
      ENDIF
      NYR=IYR(M)+1000
      IF(INDIC.EQ.0) THEN
        WRITE(6,120) IMON(M),NYR,FLOW(M),DFLOW(M),ISTAR,M,PROB(M),TIME
      ELSE
        WRITE(6,120) IMON(M),NYR,FLOW(M),DFLOW1(M),ISTAR,M,PROB(M),TIME
```

```
      ENDIF
      ILINE=ILINE+1
      PROB(M)=PROB(M)/100.
 20   CONTINUE
C
100   FORMAT('1',///,15X,'FREQUENCY ANALYSIS - LOG PEARSON TYPE III DIST
     1RIBUTION',//,12X,6A10,/)
105   FORMAT(8X,5HMONTH,5X,4HYEAR,7X,4HDATA,8X,7HORDERED,5X,4HRANK,4X,
     1 5HPROB.,3X,11HRET. PERIOD,/,8X,5H-----,5X,4H----,7X,4H----,8X.
     2 7H-------,5X,4H----,4X,5H-----,3X,11H-----------,/,9X,3H(1),7X,
     3 3H(2),8X,3H(3),10X,3H(4),8X,3H(5),5X,3H(6),8X,3H(7),/)
110   FORMAT(29X,5H(CFS),8X,5H(CFS),15X,3H(%),6X,7H(YEARS),/)
115   FORMAT(29X,5H(CMS),8X,5H(CMS),15X,3H(%),6X,7H(YEARS),/)
120   FORMAT(9X,I2,7X,I4,1X,F12.3,1X,F12.3,A1,4X,I3,4X,F5.2,5X,F8.3)
201   FORMAT(1H0,8X,I3,36H DAY LOW FLOW MEAN DISCH. IN PERIOD ,
     1 A10,A6,/)
205   FORMAT(8X,5HSTART.14X,I3,4H DAY,/,
     & 8X,5HMONTH,5X,4HYEAR,7X,4HDATA,8X,7HORDERED,5X,4HRANK,4X,
     1 5HPROB.,3X,11HRET. PERIOD,/,8X,5H-----,5X,4H----,7X,4H----,8X,
     2 7H-------,5X,4H----,4X,5H-----,3X,11H-----------,/,9X,3H(1),7X,
     3 3H(2),8X,3H(3),10X,3H(4),8X,3H(5),5X,3H(6),8X,3H(7),/)
C
      RETURN
         END
      SUBROUTINE SORTX(N,XX)
C
C  THIS S/R SORTS AN ARRAY OF FLOATING POINT VARIABLES
C  IN ASCENDING ORDER.
C
      DIMENSION XX(1)
      K=N-1
      DO 20 L=1,K
      M=N-L
      DO 20 J=1,M
      IF(XX(J)-XX(J+1)) 20,10,10
 10   XTEMP=XX(J)
      XX(J)=XX(J+1)
      XX(J+1)=XTEMP
 20   CONTINUE
      RETURN
      END
C
      SUBROUTINE GPLOT(EST,T)
      DIMENSION YARRAY(14),T(11),P(15),XARRAY(14),XARRAX(15),
     &          YARRAX(15),X1(150),Y1(150),EST(11)
      COMMON/BLOCK1/FLOW(150),PP(150),STAT(6),MN,IYR(150),IMON(150),
     &LOUT,DFLOW(150),DFLOW1(150),HOUT,INDIC,IPRD(2),MF
      COMMON/BLOCK2/QBAR,N,NET,KOUT
      J=11
      DO 50 I=1,11
 50   YARRAY(I)=EST(I)
      CALL PLOT(0.,-30.,-3)
```

```
      CALL PLOT(2.,2.,-3)
      FACT1=1.693321
      FACT2=-1.014979
      DO 100 I=1,11
      P(I)=1.0/T(I)
100   XARRAY(I)=(ALOG(-ALOG(1.0-P(I))))*FACT2+FACT1
      DO 105 I=1,J
      YARRAX(I)=YARRAY(I)
105   XARRAX(I)=XARRAY(I)
      CALL SYMBOL(0.0,6.2,0.14,STAT,0.0,60)
      CALL SCALE(YARRAY,6.0,J,1)
      IF (MN.EQ.1) CALL SYMBOL(-.37,2.35,0.14,10HFLOW (CFS),90.,10)
      IF (MN.EQ.2) CALL SYMBOL(-.37,2.35,.14,10HFLOW (CMS),90.,10)
      J1=J+1
      J2=J+2
      IF (YARRAY(J1).LT.0.) YARRAY(J1)=0.
      DO 120 I=1,7
      Y=I-1.21
      YY=YARRAY(J1)+(I-1)*YARRAY(J2)
      CALL NUMBER(-.20,Y,0.07,YY,90.0,2)
120   CONTINUE
      CALL PLOT(0.0,0.0,3)
      CALL PLOT(8.0,0.0,2)
      CALL PLOT(8.0,6.0,2)
      CALL PLOT(0.0,6.0,2)
      DO 140 I=1,5
      XI=I
      CALL PLOT(0.0,XI,3)
140   CALL PLOT(8.0,XI,2)
      DO 150 I=1,11
      IF(I.LE.3) N2=2
      IF(I.GT.5) N2=0
      IF(I.EQ.4.OR.I.EQ.5) N2=1
      IF(I.EQ.1) N2=3
      CALL PLOT(XARRAY(I),6.0,3)
      CALL PLOT(XARRAY(I),-0.00,2)
      X=XARRAY(I)-0.11
      CALL NUMBER(X,-.25,0.07,T(I),0.00,N2)
150   CONTINUE
      CALL SYMBOL(2.0,-.5,0.14,29H RECURRENCE INTERVAL IN YEARS,0.0,29)
      CALL SYMBOL(2.0,-.75,0.14,28H LOW FLOW FREQUENCY ANALYSIS,0.,28)
      YMIN=YARRAY(J1)
      YMAX=YMIN+6.*YARRAY(J2)
      N1=0
      DO 200 I=1,N
      YT=DFLOW1(I)
      IF (YT.GT.YMAX) GO TO 200
      IF (YT.LT.YMIN) GO TO 200
      XT=(ALOG(-ALOG(1.0-PP(I))))*FACT2+FACT1
      IF (XT.GT.8.) GO TO 200
      IF (XT.LT.0.) GO TO 200
      N1=N1+1
```

```
      Y1(N1)=YT
      X1(N1)=XT
200   CONTINUE
      N2=N1+1
      Y1(N2)=YARRAY(J1)
      X1(N2)=0.
      N2=N2+1
      Y1(N2)=YARRAY(J2)
      X1(N2)=1.
      CALL LINE(X1,Y1.N1,1,-1,3)
      N1=0
      DO 210 I=1,J
      YT=YARRAX(I)
      IF (YT.GT.YMAX) GO TO 210
      IF (YT.LT.YMIN) GO TO 210
      N1=N1+1
      Y1(N1)=YT
      X1(N1)=XARRAX(I)
210   CONTINUE
      N2=N1+1
      Y1(N2)=YARRAY(J1)
      X1(N2)=0.
      N2=N2+1
      Y1(N2)=YARRAY(J2)
      X1(N2)=1.
      CALL FLINE(X1,Y1,-N1,1,0,0)
      CALL SYMBOL(-1.75,-1.1,0.5,3,90.0,-1)
      CALL SYMBOL(9.25,-1.1,0.5,3,90.0,-1)
      CALL SYMBOL(9.25,7.4,0.5,3,90.0,-1)
      CALL SYMBOL(-1.75,7.4,0.5,3,90.0,-1)
      RETURN
      END
      SUBROUTINE SOL23(M,XLN,B2,B3,DB,A2,A3,IFAIL2,IFAIL3)
      REAL M,XLN(1)
      INTEGER YT,HOUT
      COMMON/BLOCK1/FLOW(150),PROB(150),STN(6),MN,IYR(150),IMON(150),LOU
     1T,DFLOW(150),DFLOW1(150),HOUT,INDIC,IPRD(2),MF
      COMMON/BLOCK2/QBAR,N,NET,KOUT
      COMMON/BLOCK6/NHA,NB,NC,THRESH,YT,UNIT(2),XNA
      KC=YT-NET
      CALL PARMS(XLN,X1,X4,X5,M)
      CALL EQN23(M,X1,XLN,X4,X5,KC,B2,A2,IFAIL2,2)
      CALL EQN23(M,X1,XLN,X4,X5,KC,B3,A3,IFAIL3,3)
      DB=B3-B2
      RETURN
      END
C
      SUBROUTINE EQN23(M,X1,XLN,X4,X5,KC,B,A,IFAIL,IQN)
C   IQN=2 IMPLIES EQN 2
C   IQN=3 IMPLIES EQN 3
      REAL M,XLN(1)
      INTEGER YT,HOUT
```

```
      COMMON/BLOCK1/FLOW(150),PROB(150),STN(6),MN,IYR(150),IMON(150),LOU
     1T,DFLOW(150),DLOW1(150),HOUT,INDIC,IPRD(2),MF
      COMMON/BLOCK2/QBAR,N,NET,KOUT
      COMMON/BLOCK6/NHA,NB,NC,THRESH,YT,UNIT(2),XNA
C  K IS THE NUMBER OF SLICES FOR SIMPSON'S RULE.
      K=100
      XC=THRESH
      IFAIL=0
      XM=X1/NET
      AX=(ALOG(XC)-M)*X5
C  SET INITIAL B LOWER AND B UPPER
C  1) LN XC> MEAN LN X
C  2) A+ OR A-; XC>M OR XC<M.
      ANN=NET
      IF ((ALOG(XC).GT.M.AND.ALOG(XC).GT.XM).OR.
     1    (ALOG(XC).LT.M.AND.ALOG(XC).LT.XM)) THEN
        IF (AX.GT.ANN) THEN
        BL=AX/(AX-NET) + 0.1
        IF (BL.LT.1.1) BL=1.1
        BU=10000.
        ELSE
        GO TO 100
        ENDIF
      ELSE
        IF (AX.GT.ANN) THEN
        BL=1.1
        BU=-AX/(ANN-AX)-0.1
        IF (BU.LT.1.1) GO TO 100
        ELSE
        BL=-AX/(ANN-AX)+0.1
        IF (BL.LT.1.1) BL=1.1
        BU=10000.
        ENDIF
      ENDIF
      IF (IQN.EQ.2)CALL EVALB2(BL,M,K,KC,XC,RL,A,XLN,X1,X4,X5)
      IF (IQN.EQ.3)CALL EVALB3(BL,M,K,KC,XC,RL,A,XLN,X1,X4,X5)
      IF (IQN.EQ.2)CALL EVALB2(BU,M,K,KC,XC,RU,A,XLN,X1,X4,X5)
      IF (IQN.EQ.3)CALL EVALB3(BU,M,K,KC,XC,RU,A,XLN,X1,X4,X5)
      IF(RU.GT..0.AND.RL.LT..0.OR.RU.LT..0.AND.RL.GT..0) THEN
C  WE HAVE A CROSS OVER, THUS BOLZANO FOR SOLUTION.
        DO 20 I=1,30
        BA=.5*(BL+BU)
        IF (IQN.EQ.2)CALL EVALB2(BA,M,K,KC,XC,RA,A,XLN,X1,X4,X5)
        IF (IQN.EQ.3)CALL EVALB3(BA,M,K,KC,XC,RA,A,XLN,X1,X4,X5)
        IF(RU.GT..0.AND.RA.LT..0.OR.RU.LT..0.AND.RA.GT..0) THEN
          RL=RA
          BL=BA
        ELSE
          RU=RA
          BU=BA
        ENDIF
20      CONTINUE
```

```
      ELSE
C NO CROSS-OVER EXISTS.
        GOTO 100
      ENDIF
C  COMPUTATION IS COMPLETE.
      B=BA
      GOTO 200
100   IFAIL=1
200   RETURN
      END
C
      SUBROUTINE PARMS(LNQ,X1,X4,X5,M)
      COMMON/BLOCK2/ QBAR,N,NET,KOUT
      REAL LNQ(1),M
C
      X1=0.
      X4=0.
      X5=0.
      DO 1 I=1,NET
      X1=X1+LNQ(I)
      X4=X4+LNQ(I)-M
      X5=X5+1./(LNQ(I)-M)
1     CONTINUE
      RETURN
      END
C
      SUBROUTINE EVALB3(B,M,K,KC,XC,X100,A,X2,X1,X4,X5)
      REAL M,X2(1)
      COMMON/BLOCK2/ QBAR,N,NET,KOUT
      A=(NET*ALOG(XC)-X1)/((ALOG(XC)-M)*(B-1.)*X5-NET*B)
      X6=0
      DO 1 I=1,NET
      X6=X6+ALOG((X2(I)-M)/A)
1     CONTINUE
C  YOU NOW HAVE A,B, AND M.
C  SUBSTITUTE KNOWN VALUES INTO EQN 4.
      YC=(ALOG(XC)-M)/A
      CALL DIGAMMA(B,X10)
      CALL SLOPEB(YC,B,X20,A,FTC)
      X100=X10*(0.-NET)+X6+KC*X20
      RETURN
      END
C
      SUBROUTINE EVALB2(B,M,K,KC,XC,X100,A,X2,X1,X4,X5)
      REAL M,X2(1),I10
      COMMON/BLOCK2/ QBAR,N,NET,KOUT
C  COMPUTE A FIRST.
      A=(NET*ALOG(XC)-X1)/((ALOG(XC)-M)*(B-1.)*X5-B*NET)
      YC=(ALOG(XC)-M)/A
C  I10=F(YC)
C  FFY=F(YC)
C  FROM THE WILSON-HILFERTY TRANSFORM.
```

```
      CALL SLPEB2(YC,B,I10,FFY,A)
C
C  SUSTITUTE KNOWN VALUES INTO EQN 2.
      X100=X4/A-B*NET-KC*(ALOG(XC)-M)/A*(FFY/I10)
100   RETURN
      END
C
      SUBROUTINE SLPEB2(YC,B,I10.FFY,A)
      REAL I10
C  YC AND B ARE TRANSFERRED FROM THE CALLING PROGRAM.
C
C  I10=F(YC)
C  FFY=F(YC)
C
      TC=3.*YC**(1./3.)*B**(1./6.)+1./(3.*B**.5)-3.*B**.5
      IF (ABS(TC).GT.20.) THEN
         FFY=0.
      I10=1.
      GO TO 100
       ENDIF
      AX=(0.17401/(1+0.33267*ABS(TC))-0.04794/(1+0.33267*ABS(TC))**2
     & +0.37393/(1+0.33267*ABS(TC))**3)*EXP(-TC**2/2)
      FTC=1.-AX
      IF (A.GT.0..AND.TC.LT.0.) FTC=AX
      IF (A.LT.0..AND.TC.GT.0.) FTC=AX
      I10=FTC
      XLNGB=ALOG(1.+1./(12.*B))+(B-.5)*ALOG(B)+.5*ALOG(2.*3.14159265)-B
      FFY=EXP(0.-YC+(B-1.)*ALOG(YC)-XLNGB)
      IF(A.LT.0.)FFY=-FFY
100   RETURN
      END
      SUBROUTINE DIGAMMA(B,X10)
C  SUBROUTINE TO COMPUTE THE DIGAMMA FUNCTION OF B.
C  X10= DIGAMMA(B)
      X10=ALOG(B+2.)-1./(2.*(B+2.))-1./(12.*(B+2.)**2)+
     1    1./(120.*(B+2.)**4)-1./(252.*(B+2.)**6.)-1./(B+1.)-1./B
      RETURN
      END
C
      SUBROUTINE SLOPEB(YC,B,X20,A,FFTC)
C  YC AND B ARE TRANSFERED FROM THE CALLING PROGRAM.
C  X20 = D/DB OF LNF(YC)
      TC=3.*YC**(1./3.)*B**(1./6.)+1./(3.*B**.5)-3.*B**.5
      IF (ABS(TC).GT.20) THEN
         X20=0.
         GO TO 100
      ENDIF
      FFTC=1/(2*3.14159265)**.5*EXP(-TC**2/2.)
      IF(A.LT.0.) FFTC=-FFTC
      AX=(0.17401/(1+0.33267*ABS(TC))-0.04794/(1+0.33267*ABS(TC))**2
     & +0.37393/(1+0.33267*ABS(TC))**3)*EXP(-TC**2/2)
      FTC=1.-AX
```

```
      IF(A.GT.0..AND.TC.LT.0.) FTC=AX
      IF(A.LT.0..AND.TC.GT.0.) FTC=AX
      X20=YC**(1./3.)/(2.*B**(5./6.))-1./(6.*B**1.5)-3./(2.*B**.5)
      X20=X20*FFTC/FTC
100   RETURN
      END
C
      SUBROUTINE GAMMA(XX,GX)
      REAL XX,GX,X,Y.GY
      IF(XX-57.) 6,6,4
C
C  THIS S/R SOLVES THE GAMMA FUNCTION FOR S/R WEIBUL
C
  4   GX=1.E38
      RETURN
  6   X=XX
      GX=1.
      IF(X-2.0) 50,50,15
 10   IF(X-2.0) 110,110,15
 15   X=X-1.
      GX=GX*X
      GO TO 10
 50   IF(X-1.) 60,120,110
 60   CONTINUE
 70   IF(X-1.) 80,80,110
 80   GX=GX/X
      X=X+1.
      GO TO 70
110   Y=X-1.
      GY= 1. + Y*(-0.5771017 + Y*(0.9858540 +Y*(-0.8764218 +
     1 Y*(0.832812 + Y*(-0.5684729 + Y*(0.2548205 + Y*(-0.05149930))))))
     2)
      GX=GX*GY
120   RETURN
      END
C
      SUBROUTINE SORTD(N,Q)
C
C  THIS S/R SORTS AN ARRAY OF PEAK FLOWS INTO DESCENDING ORDER
C
      DIMENSION Q(1)
      I=N-1
      DO 2 L=1,I
      M=N-L
      DO 2 J=1,M
      IF(Q(J)-Q(J+1)) 1,2,2
  1   QTEMP=Q(J)
      Q(J)=Q(J+1)
      Q(J+1)=QTEMP
  2   CONTINUE
      RETURN
      END
```

```
C
C
      SUBROUTINE STATS(Q,NUM,QBAR,SD,COV,G1,CK)
C
C  THIS S/R CALCULATES THE STATISTICS OF A FLOOD SERIES
C
      DIMENSION Q(1)
      SUM1=0.
      SUM2=0.
      SUM3=0.
      SUM4=0.
C
      DO 10 I=1,NUM
      SUM1=SUM1+Q(I)
 10   CONTINUE
C
      QBAR=SUM1/NUM
      DO 20 I=1,NUM
      DIFF=Q(I)-QBAR
      SQUARE=DIFF**2
      CUBE=DIFF**3
      FOUR=DIFF**4
      SUM2=SUM2+SQUARE
      SUM3=SUM3+CUBE
      SUM4=SUM4+FOUR
 20   CONTINUE
      SD=SQRT(SUM2/(NUM-1))
      THIRD=SUM3/NUM
      G1=NUM**2/(NUM-1.)/(NUM-2.)*THIRD/SD**3
      CK=((NUM*(NUM+1.))/((NUM-1.)*(NUM-2)*(NUM-3)))*(SUM4/SD**4)
      COV=SD/QBAR
      RETURN
      END
C
      SUBROUTINE ROUNDS(X,RX)
C  THIS S/R ROUNDS OFF FLOWS TO THREE SIGNIFICANT FIGURES...
      Y=X
      IF(Y.EQ.0.) GO TO 100
      IF(Y.LT.0.) X=-X
      X1=ALOG10(X)
      N1=X1
      X2=X1-N1
      X3=2.+X2
      X4=10.**X3
      N2=X4+.5
      RX=N2*10**(N1-2.)
      IF(Y.LT.0.) RX=-RX
      RETURN
100   CONTINUE
      RX=0.
      RETURN
      END
```

```
C
      SUBROUTINE LISTS
C
C  THIS S/R PRINTS OUT FLOOD DATA AND INFORMATION REQUIRED FOR A
C  HISTORICAL ANALYSIS. LOW OUTLIERS ARE NOTED BY AN ASTERISK.
C
      INTEGER YT,HOUT
      COMMON/BLOCK1/FLOW(150),PROB(150),STN(6),MN,IYR(150),IMON(150),LOU
     1T,DFLOW(150),DFLOW1(150),HOUT,INDIC,IPRD(2),MF
      COMMON/BLOCK2/QBAR,N,NET,KOUT
      COMMON/BLOCK6/NHA,NB,NC,THRESH,YT,UNIT(2),XNA
C
      NZERO=0
      ILINE=1
      DO 10 I=1,N
      IF(FLOW(I).EQ.0.0) NZERO=NZERO+1
      DFLOW(I)=FLOW(I)
 10   CONTINUE
      CALL SORTD(N,DFLOW)
C
C  HISTORICAL ANALYSIS PARAMETERS...
      NET=N-LOUT
      NA=0
      DO 12 I=1,N
      IF(FLOW(I).GE.THRESH) NA=NA+1
 12   CONTINUE
      NB=N-NA
      XNA=FLOAT(NA)
      NC=YT-N
C  WRITE OUT INFORMATION...
      WRITE(6,100) (STN(J),J=1,6)
      WRITE(6,105) YT,THRESH,UNIT(MN),N,NHA
      WRITE(6,110) NA,NB,NC
      WRITE(6,115)
      IF(MN.EQ.1) WRITE(6,130)
      IF(MN.EQ.2) WRITE(6,135)
C
C  CALCULATE ADJUSTED RANK,PROBABILITIES AND RETURN PERIODS...
      DO 20 I=1,N
      IF(DFLOW(I).LT.THRESH) ADJ=XNA+(YT-XNA)/(N-XNA)*(I-XNA)
      IF(DFLOW(I).GE.THRESH) ADJ=FLOAT(I)
      PROB(I)=(ADJ-0.4)/(YT+0.2) * 100.
      IF(DFLOW(I).EQ.0.0) PROB(I)=(XNA+(YT-XNA)/(N-XNA)*(N-NZERO-XNA)
     &)/YT*100.
      TIME=100./PROB(I)
C  PRINT OUT FLOODS,RANKS,RETURN PERIODS,ETC...
      IF(ILINE.LE.38) GO TO 15
      WRITE(6,100) (STN(J),J=1,6)
      WRITE(6,115)
      IF(MN.EQ.1) WRITE(6,130)
      IF(MN.EQ.2) WRITE(6,135)
      ILINE=1
```

```
 15   CONTINUE
      ISTAR=1H
      IF(I.GT.NET.OR.I.LE.HOUT) ISTAR=1H*
      NYR=IYR(I)+1000
      WRITE(6,120) IMON(I),NYR.FLOW(I),DFLOW(I).ISTAR,I,ADJ.
     1 PROB(I),TIME
      IF(I.EQ.NA) WRITE(6,125)
      ILINE=ILINE+1
      PROB(I)=PROB(I)/100.
 20   CONTINUE
C
100   FORMAT('1',///,8X,'HISTORICAL FLOOD FREQUENCY ANALYSIS - LOG PEARS
     1ON TYPE III DISTRIBUTION',//,15X,6A10,/)
105   FORMAT(//,10X,20HTOTAL TIME SPAN, YT=,I3.5H YRS.,8X,
     1'FLOW THRESHOLD =',F7.0,1X,A5,/,10X,'OBSERVED PEAKS, N=',
     2I3,38H  HISTORIC PEAKS ABOVE THRESHOLD, NHA=.I2,/)
110   FORMAT(23X,'OBSERVED PEAKS ABOVE THRESHOLD, NA=',I3,/,
     1 23X,'OBSERVED PEAKS BELOW THRESHOLD, NB=',I3,/,
     2 23X,' MISSING PEAKS BELOW THRESHOLD, NC=',I3)
115   FORMAT(///,10X,'MONTH YEAR      FLOOD    DESCENDING  RANK     RANK
     1       CUM.     RETURN',/,37X,'ORDER        M       ADJ.      PROB.
     2  PERIOD',/,10X,'----- ----       -----',6X,'-----          -    ---
     3-      -----      -----',/,11X,'(1)',3X,'(2)',7X,'(3)',8X,'(4)',7X,
     4'(5)',6X,'(6)',6X,'(7)',7X,'(8)',/)
120   FORMAT(11X,I2,3X,I4,1X,2(2X,F9.3),A1,I6.1X,2F9.2,F11.2)
125   FORMAT(40X,'THRESHOLD')
130   FORMAT(26X,5H(CFS),6X,5H(CFS),24X,3H(%),6X,7H(YEARS),/)
135   FORMAT(26X,5H(CMS),6X,5H(CMS),24X,3H(%),6X,7H(YEARS),/)
      RETURN
      END
      SUBROUTINE LP3HIS (LNQ,IFAIL)
      INTEGER HIST,YT,FAIL,HOUT
      REAL M,LNQ(150)
      COMMON/BLOCK1/FLOW(150),PROB(150),STN(6),MN,IYR(150),IMON(150),LOU
     1T,DFLOW(150),DFLOW1(150),HOUT,INDIC,IPRD(2),MF
      COMMON/BLOCK2/QBAR,N,NET,KOUT
      COMMON/BLOCK4/A,B,M
      COMMON/BLOCK6/NHA,NB,NC,THRESH,YT,UNIT(2),XNA
C
      IFAIL=0
      KOUT=0
146   NET=N-LOUT-KOUT
      IC=0
      IEND=0
      XF=.9
158   XMH=ALOG(XF*DFLOW(NET))
160   CALL PARMS(LNQ,X1,X4,X5,XMH)
      KC=YT-NET
      CALL EQN23(XMH,X1,LNQ,X4,X5,KC,B2,A2,IFAIL2,2)
      CALL EQN23(XMH,X1,LNQ,X4,X5,KC,B3,A3,IFAIL3,3)
      IF(IFAIL2.EQ.1.OR.IFAIL3.EQ.1)THEN
        XF=XF-.05
```

```
        IF (XF.LT..1) THEN
          IEND=IEND+1
          GOTO 180
         ELSE
          GOTO 158
        ENDIF
      ELSE
        DBH=B3-B2
      ENDIF
170   CONTINUE
C  ATTEMPT LOWER SOLUTION AT M=-4.
      XML=-4.
      CALL SOL23(XML,LNQ,B2,B3,DBL,A2,A3,IFAIL2,IFAIL3)
      IF(IFAIL2.EQ.1.OR.IFAIL3.EQ.1) GOTO 180
C  IF LANDED HERE, THEN POS SOLN OF EQNS FOUND.
C  CHECK FOR CROSS-OCER. IF NO X-OVER, THEN GO TO NEG SOLN.
      IF(DBH.GT.0..AND.DBL.LT.0..OR.DBH.LT.0..AND.DBL.GT.0.)
     1   GOTO 190
C  IF HERE, THEN NO X-OVER FOUND. THUS SOLVE FOR NEG CASE.
180   XF=1.1
181   XML=ALOG(XF*DFLOW(1))
182   CALL SOL23(XML,LNQ,B2,B3,DBL,A2,A3,IFAIL2,IFAIL3)
      IF(IFAIL2.EQ.1.OR.IFAIL3.EQ.1) THEN
        XF=XF+0.2
        IF(XF.GT.2.5) THEN
          IEND=IEND+1
          GOTO 400
        ELSE
          GOTO 181
        ENDIF
      ENDIF
C -IF YOU ARE HERE THEN A SOL FOR THE 2 EQNS JUST ABOVE MAX HAS BEEN
C  FOUND.
184   XMH=XML+10.
      CALL SOL23(XMH,LNQ,B2,B3,DBH,A2,A3,IFAIL2,IFAIL3)
      IF(IFAIL2.EQ.1.OR.IFAIL3.EQ.1) GOTO 400
C  NOTE THAT 400 IMPLIES NO SOLN FOR EITHER SIDE.
      IF(DBH.GT.0..AND.DBL.LT.0..OR.DBH.LT.0..AND.DBL.GT.0.)
     1  THEN
        GOTO 190
       ELSE
      XML=XMH
      DBL=DBH
      IF(XMH.GT.500.) GOTO 400
      GOTO 184
      ENDIF
190   DO 200 JJJ=1,15
      XMC=(XMH+XML)/2.
      CALL SOL23(XMC,LNQ,B2,B3,DBC,A2,A3,IFAIL2,IFAIL3)
      IF(IFAIL2.EQ.1.OR.IFAIL3.EQ.1) GOTO 400
      IF(DBH.EQ.0..OR.DBL.EQ.0..OR.DBC.EQ.0.) GOTO 205
      IF(DBH.GT.0..AND.DBC.LT.0..OR.DBH.LT.0..AND.DBC.GT.0.)
```

```
     1 THEN
        DBL=DBC
        XML=XMC
      ELSE
        DBH=DBC
        XMH=XMC
      ENDIF
200   CONTINUE
C  A,B,M HAVE BEEN FOUND SO WRITE THEM OUT.
205   IF(DBH.EQ.0..AND.DBC.NE.0.) THEN
        CALL SOL23(XMH,LNQ,B2,B3,DBC,A2,A3,IFAIL2,IFAIL3)
        XMC=XMH
      ENDIF
      IF(DBL.EQ.0..AND.DBC.NE.0.) THEN
        CALL SOL23(XML,LNQ,B2,B3,DBC,A2,A3,IFAIL2,IFAIL3)
        XMC=XML
      ENDIF
      GOTO 450
400   CONTINUE
      IF(KOUT.LT.10) THEN
        KOUT=KOUT+1
        GOTO 146
      ENDIF
      IFAIL=1
      GO TO 998
C
450   CONTINUE
      A=A2
      B=B2
      M=XMC
998   NET=N-LOUT
      RETURN
      END
      SUBROUTINE TRIGAM(B,X30)
C
C     X30 IS THE TRIGAMMA OF B
      X30=1./(B+2)+1./(2.*(B+2)**2)+1./(6.*(B+2)**3)
     1    -1./(30.*(B+2)**5)+1./(42.*(B+2)**7)-1./(30.*(B+2)**9)
     2    +1./(B+1)**2+1/B**2
      RETURN
      END

      SUBROUTINE VARMLC(VARA,VARB,VARM,COVAB,COVAM,COVBM)
C  PROGRAM TO COMPUTE THE COVARIANCE AND VARIANCE OF THE PARAMETERS
C FOR THE CONVENTIONAL MAXIMUM LIKELIHOOD SOLN OF THE LP3 DIST.
      REAL M
      COMMON /BLOCK2/ QBAR,N,NET,KOUT
      COMMON /BLOCK4/ A,B,M
C
      CALL TRIGAM (B,X30)
C  COMPUTE THE DETERMINANT D OF THE MATRIX.
      XN=NET
```

```
      D=(2.*X30-2./(B-1)+1./((B-1)**2))/((B-2)*A**4)
      VARA=(X30/(B-2)-1./((B-1)**2))/(XN*D*A**2)
      VARB=2./(XN*D*(B-2)*A**4)
      VARM=(B*X30-1)/(XN*D*A*A)
      COVAB=-(1./(B-2)-1./(B-1))/(XN*D*A**3)
      COVAM=(1./(B-1)-X30)/(XN*D*A*A)
      COVBM=-(B/(B-1)-1)/(XN*D*A**3)
      RETURN
      END

      SUBROUTINE SOL3HW(XX,WM,SW,CS,CVM)
C     HISTORICALLY WEIGHTED PARAMETERS.
      DIMENSION XX(150)
      INTEGER YT
      REAL SUMXNA,SUMXNB,SQ2XNA,SQ2XNB,SQ3XNA,SQ3XNB,SW2,SW,WM,CS
      COMMON /BLOCK2/ QBAR,N,NET,KOUT
      COMMON /BLOCK6/ NHA,NB,NC,THRESH,YT,UNIT(2),XNA
      NA=XNA
      NO=NET
      LOUT=NB-NO+NA
      XNB=NB
      XNY=YT
      SUMXNA=0.
      SUMXNB=0.
      SQ2XNA=0.
      SQ2XNB=0.
      SQ3XNA=0.
      SQ3XNB=0.
      IC1=NA+1
      DO 150 I=IC1,NO
  150 SUMXNB=SUMXNB+XX(I)
      DO 160 I=1,NA
  160 SUMXNA=SUMXNA+XX(I)
      W=(XNY-XNA)/XNB
      WM=(W*SUMXNB+SUMXNA)/(XNY-W*LOUT)
      DO 170 I=IC1,NO
      SQ2XNB=SQ2XNB+(XX(I)-WM)**2
  170 SQ3XNB=SQ3XNB+(XX(I)-WM)**3
      DO 180 I=1,NA
      SQ2XNA=SQ2XNA+(XX(I)-WM)**2
  180 SQ3XNA=SQ3XNA+(XX(I)-WM)**3
      SW2=(W*SQ2XNB+SQ2XNA)/(XNY-W*LOUT-1)
      SW=SQRT(SW2)
      SUMXNA=(W*SQ3XNB+SQ3XNA)*(XNY-W*LOUT)
      SUMXNB=(XNY-W*LOUT-1)*(XNY-W*LOUT-2)*SW**3
      CS=SUMXNA/SUMXNB
      CVW=SW/WM
      RETURN
      END
      SUBROUTINE EQNS(XL,AM,EM,EM1,PSII)
C
C  PURPOSE-
```

```
C     TO EVALUATE THE REDUCED MAXIMUM LIKELIHOOD ESTIMATOR FOR
C     THE LOG PEARSON TYPE III DISTRIBUTION.
C
      REAL AM,EM,EM1,PSII,Y,S1,S2,S3,S4,B,A,C,P,PSI,Q
      DIMENSION XL(150)
      COMMON /BLOCK2/ QBAR,N,NET,KOUT
      AN=NET
      S1=0.
      S2=0.
      DO 1 J=1,NET
      Y=XL(J)-AM
      S1=S1+Y
 1    S2=S2+1./Y
      B=S1*S2/(S2*S1-AN**2)
      A=S1/(AN*B)
      S3=0.
      S4=0.
      DO 2 J=1,NET
      Y=XL(J)-AM
      S4=S4+1./Y**2
 2    S3=S3+ALOG(Y/A)
      C=B+2.
      PSI=ALOG(C)-1./(2.*C)-1./(12.*C**2)+1./(120.*C**4)-1./(252.*C**6)-
     11./(1.+B)-1./B
      EM=-AN*PSI+S3
      PSII=1./C+1./(2.*C**2)+1./(6.*C**3)-1./(30.*C**5)+1./(42.*C**7)+1.
     1/(B+1.)**2+1./B**2
      P=PSII*(AN**3*S1*S4-AN**4*S2)/(S2*S1-AN**2)**2
      Q=(AN**3*S4-AN**2*S2**2)/(S2**2*S1-AN**2*S2)
      EM1=P-S2-Q
      RETURN
      END

      SUBROUTINE MAXLP3(XLN,CSL,IND)
C  PROGRAM COMPUTES PARAMETERS BY MAX LIELIHOOD.
      REAL AML,BML,MML,AMH,AMC,EMH,EML,EMC,EMH1,EML1,EMC1
      REAL PSII,EM1,EM,AM,RM,DIFF,CRIT,S1,S2,Y
      DIMENSION XLN(150)
      COMMON/BLOCK2/ QBAR,N,KN,KOUT
      COMMON/BLOCK4/ AML,BML,MML
      AN=KN
      IF(CSL) 31,31,32
 31   AMH=XLN(1)+.01
      AML=XLN(1)+50.
      GOTO 33
 32   AMH=XLN(KN)-.01
      AML=XLN(KN)-50.
 33   CALL EQNS(XLN,AMH,EMH,EMH1,PSII)
      CALL EQNS(XLN,AML,EML,EML1,PSII)
      DO 35 JJJ=1,15
      IF(EMH.GT.0..AND.EML.LT.0..OR.EMH.LT.0..AND.EML.GT.0.) GOTO 34
      GOTO 37
```

```
34    AMC=(AMH+AML)/2.
      CALL EQNS(XLN,AMC,FMC,FMC1,PSII)
      IF(FMH.GT.0..AND.FMC.LT.0..OR.FMH.LT.0..AND.FMC.GT.0.)THEN
        AML=AMC
        FML=FMC
      ELSE
        AMH=AMC
        FMH=FMC
      ENDIF
35    CONTINUE
      AM=AMH
      DO 36 I=1,50
      CALL EQNS(XLN,AM,FM,FM1,PSII)
      IF(FM1.EQ.0.) GOTO 37
      RM=AM-FM/FM1
      DIFF=ABS(RM-AM)
      CRIT=ABS(0.0001*AM)
      IF(DIFF.LT.CRIT) GOTO 38
      AM=RM
36    CONTINUE
37    CONTINUE
      IND=1
      GOTO 40
38    CONTINUE
      MML=RM
      S1=0.
      S2=0.
      DO 39 I=1,KN
      Y=XLN(I)-RM
      S1=S1+Y
39    S2=S2+1./Y
      BML=S1*S2/(S1*S2-AN**2)
      AML=S1/(AN*BML)
      CALL EQNS(XLN,MML,FM,FM1,PSII)
      IND=0
C
40    RETURN
      END

      SUBROUTINE VARMLH (VARA,VARB,VARM,COVAB,COVAM,COVBM,IT)
C-----------------------------------------------------------------
C
C  VARMLH : PROGRAM TO COMPUTE THE VARIANCE OF THE QT YEAR EVENT
C           OF A SAMPLE FROM A TRUNCATED LP3 DISTRIBUTION.
C
C  WRITTEN BY PAUL J. PILON WRB, PVM, 8TH FLOOR.
C  MAY 21, 1986.
C
C  ALL RIGHTS RESERVED.
C
C-----------------------------------------------------------------
      INTEGER YT
```

```
      COMMON/BLOCK1/FLOW(150),PROB(150),STN(6),MN,IYR(150),IMON(150),LOU
     1T,DFLOW(150),DFLOW1(150),MOUT,INDIC,IPRD(2),MF
      COMMON/BLOCK2/QBAR,N1,NET,KOUT
      COMMON/BLOCK4/A1,B1,M
      COMMON/BLOCK6/NHA,NB,NC,THRESH,YT,UNIT(2),XNA
      REAL   V(3,3),D,VARA,VARB,VARM,COVAB,COVAM,COVBM,ZT,DZDA,Z,TC,
     1       DZDB,VARZT,VARXT,SEEXT,Q,K,XC,YC,A,B,M,XM,SD,
     2       CS,T,DM2,I10,FY,DFY,EYM2,I10M2,FYM2,I10M1,FYM1,EYM1,
     3       X10,X20,I10P1,FYP1,EY,N,X30,D2TC,DTDB,DI10DB,FETC,U,XT
C  ENTER THE TRANSFORMED MEAN,STANDARD DEVIATION, AND SKEWNESS OF THE
C  GENERATED SAMPLE SO THAT A,B, AND M MAY BE COMPUTED FROM THE MOMENT
C  RELATIONSHIPS.
      A=A1
      B=B1
      NA=XNA
      NB=NB-KOUT-LOUT
      NC=NC+KOUT
      N=NA+NB
      K=NC
      YC=(ALOG(THRESH)-M)/A
C  SOLUTION OF     2
C                D LNL
C                -----
C                  2
C                DM
C
      CALL SLPEB2(YC,B,I10,FY,A)
      Q=FLOAT(NC)/FLOAT(NB+NC)
      Z=B-2.
      CALL SLPEB2(YC,Z,I10M2,FYM2,A)
      EYM2=((1.-Q)*I10M2/((B-1.)*(B-2.))+(1.-I10M2)/
     1    ((B-1.)*(B-2.)))/(1.-Q*I10)
      EYM2=N*EYM2
      DFY=FY*(B-1.)/YC - FY
      V(3,3)=(K/(A*A))*(DFY/I10-(FY/I10)**2)-(B-1.)*EYM2/(A*A)
      V(3,3)=-V(3,3)
C  SOLUTION OF     2
C                D LNL
C               -------
C                DB DM
C
      Z=B-1.
      CALL SLPEB2(YC,Z,I10M1,FYM1,A)
      EYM1=((1.-Q*I10M1)/(B-1.))/(1.-Q*I10)
      EYM1=N*EYM1
      CALL DIGAMMA(B,X10)
C    X20= D LNF(YC)
C         ---------
C         DB
C
      CALL SLOPEB(YC,B,X20,A,FETC)
      V(2,3)=-EYM1/A-K*((FY/I10)*
```

```
     1        (ALOG(YC)-X10-X20))/A
      V(2,3)=-V(2,3)
C  SOLUTION OF    2
C                D  LNL
C                ------
C                  2
C                DA
C
      Z=B+1.
      CALL SLPEB2(YC,Z,I10P1,EYP1,A)
      EY=B*(1.-Q*I10P1)/(1.-Q*I10)
      EY=N*EY
      V(1,1)=-2.*EY/(A*A)+N*B/(A*A)-K*(YC*EY/(A*I10))*(-2./A
     1       -YC*DEY/(A*EY)+YC*EY/(A*I10))
      V(1,1)=-V(1,1)
C  SOLUTION OF   2
C               D LNL
C               -----
C               DA DB
C
      DTDB=(YC**(1./3.))/(2.*(B**(5./6.)))
     1     -1./(6.*B**1.5)-1.5/(B**.5)
      V(1,2)=-N/A-K*(YC*EY/(A*I10))*(ALOG(YC)-X10-(EFTC/I10)*DTDB)
      V(1,2)=-V(1,2)
C  SOLUTION OF   2
C               D LNL
C               -----
C                 2
C               DB
C
      CALL TRIGAM(B,X30)
C
C
      D2TC=(-5.*YC**(1./3.))/(12.*B**(11./6.))+.25/(B**2.5)+
     1     .75/(B**1.5)
C     DLN=DLN(DF(YC)/DB)/DB
      TC=(YC**(1./3.)+1./(9.*B**(2./3.))-B**(1./3.))*3*B**(1./6.)
      DLN=D2TC/DTDB - DTDB*TC
      DI10DB=EFTC*(YC**(1./3.)/(2.*B**(5./6.))-1./(6.*B**1.5)
     1       -1.5/B**.5)
C
      V(2,2)=-N*X30+K*DI10DB*(DLN-DI10DB/I10)/I10
      V(2,2)=-V(2,2)
C  SOLUTION OF   2
C               D LNL
C               -----
C               DM DA
C
      V(1,3)=-N/(A*A)+K*(YC*DEY/(A*A*I10)-((EY/I10)**2)*
     1       YC/(A*A)+EY/(A*A*I10))
      V(1,3)=-V(1,3)
C  NOW FILL IN THE REST OF THE MATRIX
```

```
      V(3,1)=V(1,3)
      V(2,1)=V(1,2)
      V(3,2)=V(2,3)
C
C  THE INVERSE DISPERSION MATRIX IS NOW COMPLETE.
C  COMPUTE THE DETERMINANT OF THE MATRIX.
      IT=0
5     D=V(1,1)*V(2,2)*V(3,3)-V(1,1)*V(2,3)*V(3,2)-V(1,2)*V(2,1)*
     1  V(3,3)+V(1,2)*V(2,3)*V(3,1)+V(1,3)*V(2,1)*V(3,2)-
     2  V(1,3)*V(2,2)*V(3,1)
      IF(D.LE.0.) THEN
        IT=IT+1
        IF(IT.LT.5) THEN
          FACTOR=1.00005
        ELSE
          FACTOR=1.001
        ENDIF
        V(2,2)=V(2,2)*FACTOR
        IF(IT.LT.9) THEN
          GOTO 5
        ELSE
          IT=9
          GOTO 1000
        ENDIF
      ENDIF

C
      VARA=(V(2,2)*V(3,3)-V(3,2)**2)/D
      VARB=(V(1,1)*V(3,3)-V(1,3)**2)/D
      VARM=(V(1,1)*V(2,2)-V(1,2)**2)/D
C
      COVAB=-1*(V(1,2)*V(3,3)-V(1,3)*V(3,2))/D
      COVAM=(V(2,1)*V(3,2)-V(3,1)*V(2,2))/D
      COVBM=-1*(V(1,1)*V(3,2)-V(3,1)*V(1,2))/D
1000  RETURN
      END
      SUBROUTINE FPLOT(EST,T,HIST,FAIL)
C
C     PURPOSE -
C        TO PLOT A FREQUENCY CURVE FITTED USING THE LOG PEARSON
C        TYPE III DISTRIBUTION
C
C     DESCRIPTION OF VARIABLES
C        YARRAY ....DESIGN FLOOD ESTIMATES TO BE PLOTTED
C        N      ....SAMPLE SIZE
C        T      ....RETURN PERIODS OF DESIGN FLOODS
C        STN    ....STATION NUMBER AND NAME
C        DFLOW  ....OBSERVED FLOODS IN DESCENDING ORDER OF MAGNITUDE
C        PROB   ....PROBABILITIES OF OBSERVED FLOODS
C
      INTEGER HIST,FAIL
      COMMON/BLOCK1/FLOW(150),PROB(150),STN(6),MN,IYR(150),IMON(150),
```

```
     &LOUT,DFLOW(150),DFLOW1(150),HOUT,INDIC,IPRD(2),mF
      COMMON/BLOCK2/QBAR,N,NET,KOUT
      DIMENSION YARRAY(15),P(15),XARRAY(15),X1(150),XP(15),T(1),
     &XFLOW(150),YY(15),EST(1),ZARRAY(15)
      DATA(XP(I),I=1,11)/-2.750,-1.645,-0.842,0.0,0.842,1.282,1.645,2.05
     14,2.326,2.575,2.88/
C
      CALL PLOT(0.,-30.,-3)
      CALL PLOT(2.,2.,-3)
      FACT1=1.46628
      FACT2=3.77713
      DO 1 I=1,11
      XARRAY(I)=XP(I)*FACT1+FACT2
      P(I)=1./T(I)
1     CONTINUE
      II=0
      DO 2 I=1,11
      IF(EST(I).LE.0.001) GO TO 2
      II=II+1
      YARRAY(II)=EST(I)
2     CONTINUE
      DO 3 I=1,11
      IF (I.LE.3) N2=2
      IF (I.GT.5) N2=0
      IF (I.EQ.4.OR.I.EQ.5) N2=1
      IF (I.EQ.1) N2=3
      X=XARRAY(I)-0.10
      CALL PLOT (XARRAY(I),6.0,3)
      CALL PLOT (XARRAY(I),0.0,2)
      CALL NUMBER (X,-.25,0.07,T(I),0.0,N2)
3     CONTINUE
C  REDEFINE N IF ZERO FLOWS ARE PRESENT IN OBSERVED FLOW SERIES...
      NUM=0
      DO 21 L=1,N
      IF(DFLOW(L).LT.0.001) GO TO 21
      NUM=NUM+1
 21   CONTINUE
      N=NUM
      DO 4 III=1,11
      I=III
      XTEMP=10.**(I-4)
      IF(XTEMP.GT.YARRAY(1).OR.XTEMP.GT.DFLOW(N)) GO TO 5
4     CONTINUE
    5 III=II+1
      II2=II+2
      YARRAY(II1)=10.0**(I-5)
      I=I-5
      DO 6 III=1,140
      K=III
      XTEMP=10.**K
      IF(XTEMP.GT.YARRAY(II).AND.XTEMP.GT.DFLOW(1)) GO TO 7
6     CONTINUE
```

```
 7 YARRAY(II2)=(K-I)/6.0
   DO 10 K=1,70
   IF (K.LE.10) I=K
   IF (K.LE.19.AND.K.GT.10) I=I+10
   IF (K.LE.28.AND.K.GT.19) I=I+100
   IF (K.LE.37.AND.K.GT.28) I=I+1000
   IF (K.LE.46.AND.K.GT.37) I=I+10000
   IF (K.LE.55.AND.K.GT.46) I=I+100000
   IF (K.LE.64.AND.K.GT.55) I=I+1000000
   XH=ALOG10(YARRAY(II1)*I)/YARRAY(II2)-ALOG10(YARRAY(II1))/YARRAY(II
  *2)
   IF (XH.GT.6.001) GO TO 11
   CALL PLOT (-0.25,XH,3)
   CALL PLOT (8.0,XH,2)
   IF (K.EQ.1.OR.K.EQ.19.OR.K.EQ.28.OR.K.EQ.37.OR.K.EQ.46.OR.K.EQ.10.
  *OR.K.EQ.55.OR.K.EQ.64) GO TO 8
   XI=XI+1.
   X3=XH-0.035
   CALL NUMBER (-0.30,X3,0.07,XI,90.0,-1)
   GO TO 9
8  CONTINUE
   XI=1.
   XI1=10.
   X3=XH+0.02
   XI2=ALOG10(YARRAY(II1)*I)
   CALL NUMBER (-0.30,X3,0.10,XI1,90.0,-1)
   X2=X3+0.18
   CALL NUMBER (-0.37,X2,0.07,XI2,90.0,-1)
9  CONTINUE
10 CONTINUE
11 CONTINUE
   XFLOW(N+1)=0.
   XFLOW(N+2)=1.
   DFLOW(N+1)=YARRAY(II1)
   DFLOW(N+2)=YARRAY(II2)
   X1(N+1)=0.0
   X1(N+2)=1.0
   XARRAY(12)=0.0
   XARRAY(13)=1.0
   DO 14 I=1,N
   IF (PROB(I).LT.0.5) GO TO 12
   ICHECK=0
   P1=1.0-PROB(I)
   GO TO 13
12 ICHECK=1
   P1=PROB(I)
13 T1=(ALOG(1.0/P1**2))**0.5
   XP1=T1-(2.51552+0.80285*T1+0.01033*T1**2)/(1.0+1.43279*T1+0.18927*
  1T1**2+0.001318*T1**3)
   IF (ICHECK.EQ.0)  X1(I)=FACT2-XP1*FACT1
   IF (ICHECK.EQ.1)  X1(I)=FACT2+XP1*FACT1
   XFLOW(I)=ALOG10(DFLOW(I))/YARRAY(II2)-ALOG10(YARRAY(II1))/YARRAY(I
```

```
     &I2)
14    CONTINUE
      CALL LINE (X1,XFLOW,N,1,-1,3)
      DO 15 I=1,II
15    YY(I)=ALOG10(YARRAY(I))/YARRAY(II2)-ALOG10(YARRAY(II1))/YARRAY(II2
     *)
      YY(II1)=0.
      YY(II2)=1.
      JJ=12-II
      MM=0
      DO 16 I=JJ,13
      MM=MM+1
      ZARRAY(MM)=XARRAY(I)
16    CONTINUE
      CALL FLINE(ZARRAY,YY,-II,1,0,0)
      CALL SYMBOL (0.30,6.6,0.14,STN,0.0,55)
      IF(MN.EQ.1) CALL SYMBOL(-0.5.1.88,0.14,15HDISCHARGE (CFS),90.,15)
      IF(MN.EQ.2) CALL SYMBOL(-0.5,1.88,0.14,15HDISCHARGE (CMS),90.,15)
      IF(HIST.EQ.0) CALL SYMBOL(1.40,6.4,0.10,51HFLOOD FREQUENCY - LOG P
     1EARSON TYPE III DISTRIBUTION,0.0,51)
      IF(HIST.GT.0) CALL SYMBOL(1.00,6.4,0.10,62HHISTORICAL FLOOD FREQUE
     1NCY - LOG PEARSON TYPE III DISTRIBUTION,0.0,62)
      IF(FAIL.LT.1) CALL SYMBOL(2.53,6.2,0.07,42HPARAMETERS ESTIMATED BY
     1 MAXIMUM LIKELIHOOD,0.0,42)
      IF(FAIL.GT.0) CALL SYMBOL(2.94,6.2,0.07,31HPARAMETERS ESTIMATED BY
     1 MOMENTS,0.0,31)
      CALL SYMBOL (2.04,-.6,0.14,28HRECURRENCE INTERVAL IN YEARS,0.0,28)
      CALL SYMBOL (-1.75,-1.1,0.5,3,90.0,-1)
      CALL SYMBOL (9.25,-1.1,0.5,3,90.0,-1)
      CALL SYMBOL (9.25,7.4,0.5,3,90.0,-1)
      CALL SYMBOL (-1.75,7.4,0.5,3,90.0,-1)
      RETURN
      END
```