Utah State University

DigitalCommons@USU

Computer Program Generation of Extreme Value Distribution Data

Stephen (Wan-tsing) Lei
Utah State University

Utah State University
MERRILL-CAZIER LIBRARY

COMPUTER PROGRAM GENERATION OF EXTREME VALUE

DISTRIBUTION DATA

by

Stephen (Wan-tsing) Lei

A report submitted in partial fulfillment
of the requirements for the degree

of

MASTER OF SCIENCE

in

Applied Statistics

Approved:

UTAH STATE UNIVERSITY
Logan, Utah
1986

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURE

ABSTRACT

COMPUTER PROGRM GENERATION OF EXTREME VALUE

DISTRIBUTION DATA

by

Stephen (Wan-Tsing) Lei, Master of Science

Utah State University, 1986

Major Professor: Dr. Ronald V. Canfield
Department: Applied Statistics

The application of the Monte Carlo method on the
estimation in Gumbel extreme value distribution was
studied. The Gumbel extreme value distribution is used
to estimate the flood flow of specific return period
for the design of flood mitigation project. This paper
is a programming effort (1) to estimate the parameters
of Gumbel distribution using the observed data and (2)
to provide a random variate generating subroutine
to generate random samples and order statistics of
a Gumbel distribution random variable. The mean
squared error is used to measure the accuracy of
the estimation method. Finally, an example of the
use of these programs is given to illustrate application
in the analysis of a hydrologic system.

(29 pages)

CHAPTER I

INTRODUCTION

An important tool in statistical analysis is the capability of generating random variates on the computer. Generated data is used in Monte Carlo anaylsis of complex problems which cannot be solved analytically. The Gumbel distribution is an important distribution in hydrology. It is used to decribe the distribution of extreme events such as yearly floods and droughts. This paper is a programming effort (1) to estimate the parameters of Gumbel distribution using the observed data and (2) to provide a random variate generating subroutine to generate random samples and ordered statistics of a Gumbel distributed random variable. An example of the use of these programs is given to illustrate application in flood hydrology.

Relevance of Investigation

With the continuing development of flood plains and rural watersheds for urban use, flood control becomes increasingly important. Construction of dams, water needed for irrigational purposes, keeping a river within its embankment, all require estimation of flood frequency and severity, through statistical analysis of the extreme value hydrologic flood flow data.

The design of structures related to water resources management and control is heavily dependent on the extreme hydrologic event. Design parameters usually include the yearly maximum event with n - year return period. Prediction of flood frequency by height is the basis of most of the specifications for flood control. Considerable effort has been expended to determine a distribution of maximum yearly river height which can be applied uniformly to all streams with reasonable accuracy. After fitting several distributions to many different data sets representing a wide variety of conditions, the log Pearson Type III distributions have been judged to give the best overall fit (Benson, 1968). This has been reinforced by the work of Beard (1974). There has been disagreement that the log Pearson Type III distributions are "best" (Bobee and Robitaille, 1977), and Reich (1977) has questioned the advisability of even suggesting that uniform method may apply.

Selecting a distribution to describe floods has been essentially one of curve fitting. It is very necessary in the application of these distributions for design and management decisions to extrapolate, i.e. to estimate return periods beyond the range of the data. Thus the hydrologist is forced to make decisions in regions in which he has no data. A serious difficulty is inherent when one uses empirical fit to select a distribution for maximum river

height.   Many different distributions can provide a good
empirical fit in the range of the data and yet have very
different right tail characteristics.   The most important
consideration in selecting a distribution for use   in
describing   maximum yearly river height is the behavior of
the right tail of the distribution. It is from the right tail
that return periods and probabilities of rare events are
determined.   Considering the region where greatest accuracy
is needed, empirical fit of a distribution over the data set
is   not   adequate as sole criterion for choosing   a
distribution.   Some theoretical principle is needed to
assist in the choice of a distribution due to the absence of
data in the right tail.   The application of extreme value
distributions to hydrology have been studied by Canfield et,
al, (1980).   When used properly they are shown to have both
theoretical basis and good empirical fit.

Objective of study

Every year, floods cause loss of life and millions of
dollars worth of damage around the United States , for
example, in April 1983 there was a spring flood causing the
damage over 200 millions dollars in Utah.   How to mitigate
the damage economically can provide the way to use the water
resources and to manage the flood alluvial fan , Wasatch
Front, in Utah.   The flood flow frequency analysis is a
tool which is used to reach that goal.

Frequency analysis is used not only as an aid in averting disaster but is also a means of introducing efficient flood control structural and non - structural designs. Anaylsis of such structures is often very complex requiring Monte Carlo methods. In order to generate data with a given distribution it is necessary to define all parameter values. Therefore this work is divided into two main parts. A program for estimation of the parameters of the Gumbel distribution is written and tested in the first part. Generation of Gumbel data both random and ordered is considered in the second part. The mean squared error is used to measure the accuracy of the estimation method. The final chapter of this work is devoted to an example which serves to illustrate the use of these programs in the analysis of a hydrologic system for flood flow estimation.

CHATPER II

LITERATURE REVIEW

Data Sampling

The available hydrologic data are generally presented in chronological order. For extreme value distribution sampling there are two ways; annual maximum series and annual exceedance series. Chow (1964) states the annual exceedance series should be used for designing a bridge foundation because flooding sometimes results from the repetition of flood occurrance rather than from a single peak flow. In other cases where the design is governed by the most critical condition, such as spillway design, the annual maximum series should be used. Laugbein (1949) investigated the relationship between the probabilities of the annual exceedance series and annual maximum series and found that the difference between them is not very significant except in the value at low river flow.

Extreme Value Theory

L. Tippett (1925) calculated the probability of largest normal values for different sample sizes up to 1000, and the mean normal range for samples for two to 1000. His

table become the fundamental tool for all practical uses of largest values from normal distributions. The first study of largest value for other distributions was made by E. L. Dodd (1923). His work is based on "asymptotic" values which are again similar to the characteristic largest value. Frechet, M. (1927) showed that largest values taken from different initial distributions sharing a common property may have a common asymptotic distribution. He introduced the stability postulate according to which the distributions of the largest value should be equal to the initial one, except for a linear transformation. R. A. Fisher and L. H. Tippett(1928) used the same stability postulate, and found in addition to Frechets' asymptotic distributions, two others valid for other initial types. R. von. Mises (1936) classified the initial distribution's possessing asymptotic distributions of the largest value, and gave sufficient conditions under which the three asymptotic distribtuions are valid. J. E. Gumbel (1958) worked on the return period of flood flow. He derived the type I extreme value distribution equation as follows:

1. Let $x_1, x_2, \ldots, x_N$ be a series of independent random variables with cumulative probability distribution given by:

$$P(X)=P(x_V \leq y) \tag{2-1}$$

2. Define X_N as the maximum value of x in a sample of

length N, i.e. $X_N = \underset{1 \leq v \leq N}{MAX} \; x_v$ so that

$$P(X_N \leq y) = P(x_1 \leq y, x_2 \leq y, \ldots x_N \leq y)$$

or $P(X_N \leq y) = [P(y)]^N$ (2-2)

3. Now assume that the tail of the distribution P(y) is exponential such that

$$P(y) = 1 - a\, e^{-y} \quad (2\text{-}3)$$

4. From equation (2-2) if Ln(aN) is a normalizing constant

$$P(X_N \leq y + Ln(aN)) = [P(y + Ln(aN))]^N \quad (2\text{-}4)$$

and from Equation (2-3)

$$P(y + Ln(aN)) = 1 - ae^{-(y + Ln(aN))} \quad (2\text{-}5)$$

so that

$$P(X_N \leq y + Ln(aN)) = [1 - ae^{-(y + Ln(aN))}]^N$$

or

$$P(X_N \leq y + Ln(aN)) = [1 - e^{-y}/N]^N \quad (2\text{-}6)$$

5. If $N \to \infty$, then:

$$\lim_{N \to \infty} P(X_N \leq y + Ln(aN)) = \lim_{N \to \infty} [1 - e^{-y}/N]^N$$

or

$$\lim_{N \to \infty} P(x_N \leq y + Ln(an)) = \exp(-e^{-y}) \quad (2\text{-}7)$$

$$P(x) = \exp(-e^{-y})$$
or

$$P(x) = \exp(-e^{-a(X + B)}) \quad (2\text{-}8)$$

Parameters Estimation

1. Moment Method:

Fisher and Tippett(1928) applied the standard equation for generation of original moments to the pdf of the reduced type I extreme value distribution as follows:

$$U'_{y,r} = \int_{-\infty}^{\infty} y^r \exp(-y-e)^{-y} dy \qquad (2\text{-}9)$$

Substituting z for e^{-y} the result is

$$U'_{y,r} = \int_{-\infty}^{\infty} (-\ln z)^r z e^{-z} (-dz/z) \qquad (2\text{-}10)$$

The first moment about the original, r=1, is then

$$U'_{y,1} = \int_{o}^{\infty} \ln z \; (z) \; e^{-z} \; dz/z \qquad (2\text{-}11)$$

But $\int_{o}^{\infty} z e^{-z} dz$ is $\lceil(1)\rceil$, so that, from Kendall and Stuart (1958)

$$U'_{y,1} = -\psi(1) = \gamma$$

where γ is Euler's constant, approximately 0.57721. Reconverting to the original variate x as x = y/a + B

$$U'_1 = B + \gamma/a \qquad (2\text{-}12)$$

Similarly, Gumbel (1958) has shown that second moment about the mean, u_2 , is given by

$$u_2 = \pi^2/(6a^2) , \qquad \sigma = \pi/(\sqrt{6}\,a)$$

$$a = \pi/(\sqrt{6}\,\sigma) = 1.2825/\sigma \qquad (2\text{-}13)$$

$$B = u - \gamma/a = u - (0.57721/a) \qquad (2\text{-}14)$$

"B" is the mode of the distribution or location parameter.

"a" is the distribution dispersion parameter or shape parameter.

2. Weighted Least Squares Method:

A weighted least squares technique reported by Bain

and Antle (1967) and the improvement of this method by White (1969) is shown as follows:

Gumbel extreme value equation,

$$P(x) = \exp(-e^{-Y}) \tag{2-15}$$

Take the double log of $P(x)$, then

$$a\ (X_{in}-B) = \log\ (\ -\log\ P(x)) = Y_i \tag{2-16}$$

$$Y_i = f(\ a,\quad B,\quad X_{in}) \tag{2-17}$$

$$Z = \Sigma\ (E_i - Y_i)^2 W_i \tag{2-18}$$

Thus

$$Z = \Sigma(\ a\ X_{in} - a\ B - E_i)^2 W_i \tag{2-19}$$

$$Q(X_i) = X_i \tag{2-20}$$

where

E_i is known value, $E_i = E(Y_i)$

W_i is the wieght function of Y_i, $W_i = [Var(Y_i)]^{-1}$

X_{in} is order statistic.

Kwan (1979) compared between the weighted least square method and the maximum likelihood method for estimating the parameters of type I extreme value distribution. The estimating formula of weighted least squares method are listed below, comparing with the Weibull distribution analysis,

$$Z = \Sigma\ [h_1(a,B)+h_2(a,B)Q(X_{in})-E_i]^2 W_i \tag{2-21}$$

$$\widehat{h_2(a,B)} = \frac{\Sigma(Q(X_{in})-\bar{Q})(E_i-\bar{E})W_i}{\Sigma(Q(X_{in})-\bar{Q})^2 W_i} \tag{2-22}$$

$$\widehat{h_1(a,B)} = \bar{E} = h_2(a,B) \ \bar{Q} \tag{2-23}$$

$$\bar{E} = \frac{\Sigma E_i \ W_i}{\Sigma W_i} \tag{2-24}$$

$$\bar{Q} = \frac{\Sigma Q(X_{in}) \ W_i}{\Sigma W_i} \tag{2-25}$$

comparing equation (2-19) with equation (2-21)

$$h_1(a \ ,B) = -aB$$

$$h_2(a \ ,B) = a$$

$$Q(X_{in}) = X_{in}$$

Then $a = h_2(a,B)$ and $B = -h_1(a,B)/a$

3. Maximum Likelihood Method :

The maximum likelihood estimation of the type I extreme value distribution was first proposed by Kimball (1946). It is not practical until the advent of computers. Harter and Moore(1968) investigated the maximum likelihood method for estimating the type I extreme value distribution, but the numerical procedure requires considerable computer time. Panchang (1967) reported a more efficient numerical procedure for the maximum likelihood estimation of the parameters of type I extreme value distribution. This method is widely used in computer programming. Details of the numerical procedure are shown in

chapter III.

Monte Carlo Method

Nash and Armorocho(1966)developed expressions for the standard error of sample estimates for flood magnitudes of a specified return period, using Monte Carlo sampling technique from a double exponential distribution, and assumed Gumbel straight line. Canfield (1980)used a Monte Carlo experiment to apply extreme value theory in estimating flood peaks from mixed populations. A five parameter distribution was applied to 11 long-term sequences and shown by the plotting test to originate from nonhomogeneous sources. The fit was generally excellent. Wallis et al (1985) uses the generalized extreme value distribution which combines three possible extreme value distributions to analyze the flood frequency.

CHAPTER III

METHODOLOGY

The first section is using the observed flood data to estimate the Gumbel distribution parameters, A and B by applying maximum likelihood estimation method. The Panchang (1967) numerical procedure will be stated in detail. In the second section the estimated parameters, A and B are used in the generate random samples from the Gumbel distribution. The Monte Carlo method is explained step by step. The computer programs are listed in the appendix. The final section is to measure the accuracy of the estimation method by using mean squared error.

Maximum Likelihood Estimation

For a continuous distribution, the likelihood $L(Q)$ for a complete sample of n observations $y_1, \ldots \ldots y_N$ is defined as the joint probability density $f(y_1, \ldots y_N; Q)$. The likelihood $L(Q)$ is viewed as a function, an arbitrary value of the distribution parameter. The true value (unknown) is denoted by Q_o. Usually the N observations are a random sample of independent observations from the same distribution with probability density $f(Y, Q)$. Then the sample likelihood is,

$$L(Q) = f(Y_1; Q) \ f(Y_2; Q) \ldots \ldots f(Y_n; Q) \qquad (3-1)$$

The method of maximum likelihood is to estimate, Q, such that L(Q) is maximized. This is obtained by partially differentiating L(Q) with respect to each of the parameters and equating to zero.

Frequently Ln L(Q) is used instead of L(Q) to simplify computations. The following is the maximum likelihood method for evaluating the Gumbel distribution parameters. If the probability of X_1 occurring as an annual highest peak, it can be expressed as follows,

$$f(X_1) = ae^{-a(X_1-B)}Exp(-e^{-a(X_1-B)}) \qquad (3-2)$$

same as for X_2

$$f(X_2) = ae^{-a(X_2-B)}Exp(-e^{-a(X_2-B)}) \qquad (3-3)$$

In general terms

$$f(X_1,..X_N) = a^N e^{-\Sigma a(X_i-B)}Exp(-e^{-\Sigma a(X_i-B)}) \qquad (3-4)$$

$$L=\log f(X_1,..X_N) = N\log_e a - \Sigma a(X_i-B) - \Sigma e^{-a(X_i-B)} \quad ..(3-5)$$

Then maximize the likelihood function as follows;

$$\partial L/\partial B = 0 \quad and \; \partial L/\partial a = 0 \qquad (3-6)$$

$$B=(\log_e 10 \; \log_{10}N)/a-(\log_e 10 \; \log_{10}(\Sigma e^{-aX}i))/a \qquad (3-7)$$

$$F(a) = \Sigma X_i e^{-aX_i} - (\bar{X}-1/a) \Sigma e^{-aX_i}=0 \qquad (3-8)$$

Panchang (1967) used a Taylor series expansion to solve equation (3-8) as follows;

$$F'(a)=\frac{dF(a)}{d(a)}$$

$$=-\Sigma X_i^2 e^{-aX_i}+(\bar{X}-1/a)\Sigma X_i e^{-aX_i}-(1/a)^2\Sigma e^{-aX_i} \qquad (3-9)$$

Panchang(1967) used the sucessive approximations to estimate

a_1 by applying the following equation.

$$h_1 = \frac{-F(a_1)}{F'(a_1)} \tag{3-10}$$

then, $F(a_2) = F(a_1+h_1) = F(a_1)+h_1F'(a_1)$ (3-11)

and $\qquad\qquad a_2 = a_1 + h_1$ (3-12)

This procedure is repeated until a sufficiently small value of $F(a_1)$ is obtained when B can be obtained from equation (3-7). Generally, only 3 or 4 steps will be required.

Monte Carlo Experiment

The problem encountered when empirical fit is the sole criterion used to select a "best" distribution to describe a population increases as one uses the distribution to estimate the frequency of rare events. It is sometimes suggested that no distribution is perfect; therefore, several may do an adequate job, and certainly the "best" fit will be close. This argument may be valid when the distributions are used to estimate probabilities of return periods for frequently occurring events. However, when estimates are needed for extreme or rare events, serious errors can result from use of a distribution selected on the basis of empirical fit because the probabilities of rare events are computed from the tails of a distribution, whereas empirical fit is dominated by the body of the data set. Therefore the Monte Carlo method is used to generate

more data for evaluating the probability of rare events. The steps of the Monte Carlo experiment were shown as follows:

1. From a known distribution (or probability law) a sequence of psuedo-random numbers was generated (equivalent to a single hypothetical flood record for a site).

2. Using the psuedo-random numbers to generate a random variate.

3. From a known distribution (Gumbel distribtution) parameters were estimated using the maximum likelihood method.

4. The parameters of the generating distribution were compared with the estimated parameters of the observed samples.

5. Steps 1-4 were repeated many times.

6. Steps 1-5 is refered as a Monte Carlo experiment.

In this paper, samples of size 59, were generated 30, 50, 100, and 500 times from a Gumbel population. The detailed steps were shown as follows:

1. Using Fortran 77, on a VAX 11/780 computer, the uniform random number generator was used to generate random numbers.

2. The inverse transformation technique was used to generate the Gumbel variate. The procedure is as follows:

a). estimate the CDF of the derived random variable X.

b). set F (X) = U on the range of X.

c). solve the equation F(X)=U for X as a function of U.

d). Repeat steps a - c as needed.

For the Gumbel distribution:

a). $F(X) = Exp(-e^{-a(X-B)})$ (3-13)

b). $Ln\ F(X) = -e^{-a(X-B)} = Ln(U)$ (3-14)

c). $Ln[-LnF(X)] = -a(X-B) = Ln[-Ln(U)]$ (3-15)

d). $X = -Ln[-Ln(U)]/a + B$ (3-16)

The Fortran Statement for generating random variate, is programmed in subroutine RANDOM as

X(I) = B - DLOG[-DLOG(U)]/a

where U is a uniform random number generated and ordered by the following statement,

U = 1.0 - (1.0-TEMP)*Y**(1.0/(FLOAT(N-I)+1.0))

Model Accuracy Test

The following formulae was used to calculate the accuracy of the model,

$$\text{Sum of Square Error} = \frac{\sum \left[\begin{array}{c} \text{Esti. A by using} \\ \text{generating Sample} \end{array} - \begin{array}{c} \text{Esti. A by using} \\ \text{observed Sample} \end{array} \right]^2}{\text{Number of Monte Carlo Trials}} \qquad (3-17)$$

$$\text{Sum of Square Error} = \frac{\sum \left(\begin{array}{l} \text{Esti. B by using} \\ \text{generating Sample} \end{array} - \begin{array}{l} \text{Esti. B by using} \\ \text{observed Sample} \end{array} \right)^2}{\text{Number of Monte Carlo Trials}} \qquad (3\text{-}18)$$

$$\text{Average} = \frac{\sum \text{Estimator A}}{\text{Number of Monte Carlo Trials}} \qquad (3\text{-}19)$$

$$\text{Average} = \frac{\sum \text{Estimator B}}{\text{Number of Monte Carlo Trials}} \qquad (3\text{-}20)$$

The flow chart of the Monte Carlo program is shown on Figure 1.
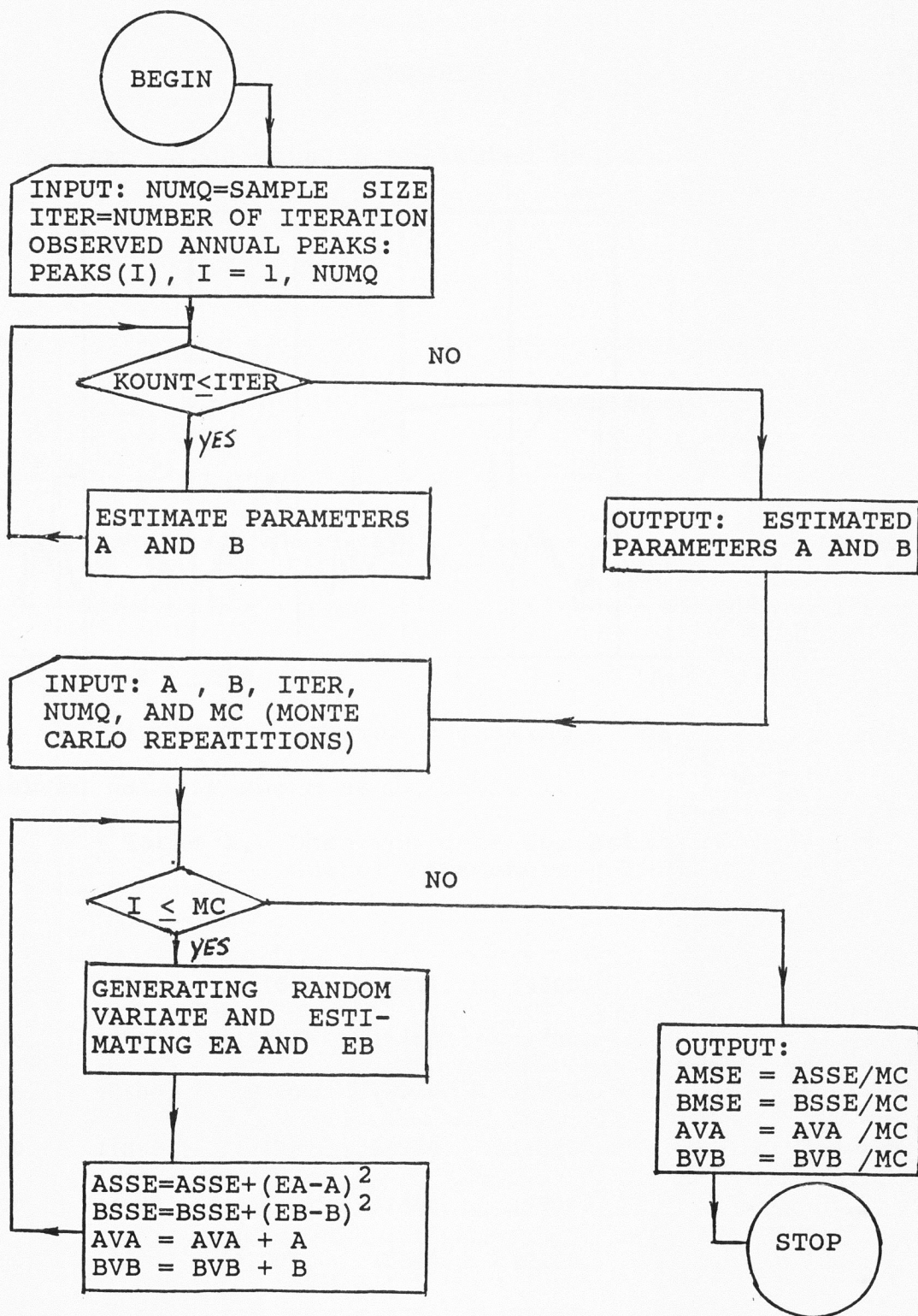
Figure 1. The Flow Chart of the Program

CHAPTER IV

EXAMPLE

Extreme value data is collected as   shown   in Figure 2.
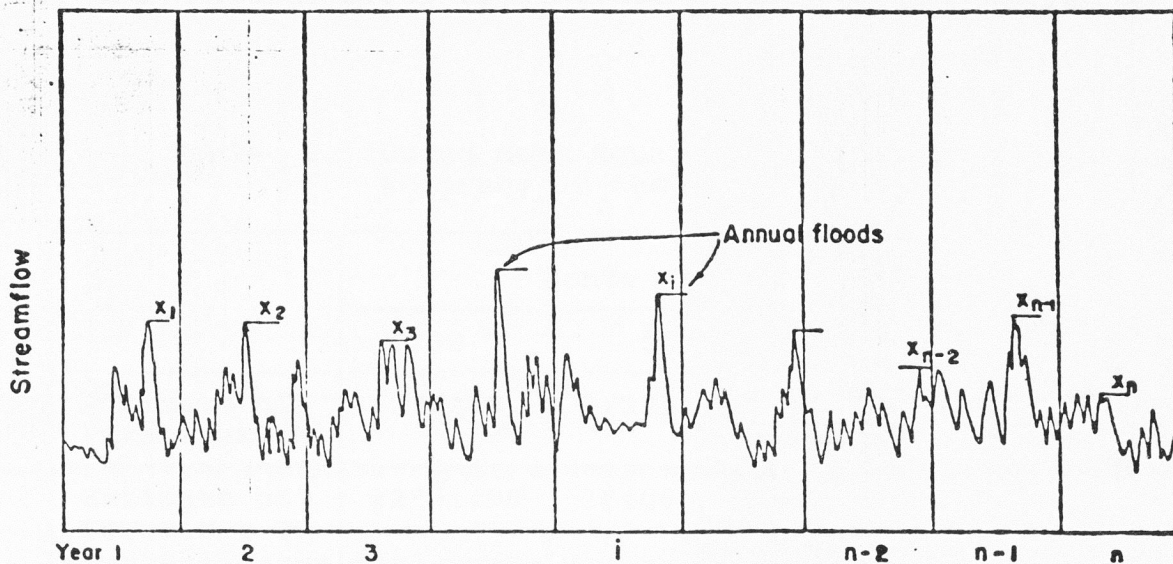
Figure 2.   Schematic showing   data collection

Observed data is shown as in Table 1.

Table 1.   Observed data for Estimating the
Gumbel Parameters

59, 4

10040.	10700.	20100.	8210.	14300.	8040.	8210.	1390
8390.	18500.	13000.	16400.	14500.	13000.	17200.	1390
11900.	13600.	12400.	18300.	12900.	18200.	9900.	1020
9020.	11800.	16100.	16900.	11800.	13900.	12300.	1510
11900.	11000.	16000.	11600.	19900.	18600.	18000.	1310
29100.	10300.	12200.	15600.	12700.	13100.	19200.	1950
23000.	6700.	7130.	14300.	20600.	25600.	8180.	3440
16100.	10200.	12300.					

Results: Using the observed data,

$$\text{Estimator} \quad a = 0.2375*10^{-3}$$

$$\text{Estimator} \quad B = 12156.98$$

The accuracy of the estimated a and B is measured by mean squared error testing method (as shown in Table 2.)

Table 2. Using Mean Squared Error to Measure the Accuracy of the Monte Carlo Method

	Monte	Carlo	Number	
	30	50	100	500
Estimate of Parameter EA	$.237*10^{-3}$	$.237*10^{-3}$	$.237*10^{-3}$	$.237*10^{-3}$
Estimate of Parameter EB	12561.98	12400.12	12278.55	12181.29
Mean Squared Error of a, slope para.	$.668*10^{-12}$	$.1098*10^{-12}$	$.495*10^{-12}$	$.886*10^{-12}$
Mean square error of B, location para.	385.01	4707.85	6.07	75.80
Bias in Estimator a, shape para.	$.791*10^{-5}$	$.475*10^{-5}$	$.237*10^{-5}$	$.475*10^{-6}$
Bias in Estimator B, location para.	405.23	243,14	121,57	24.31

REFERENCES

Bain, L. J. and Antle, C. E., 1967. "Estimation of Parameters in the Weibull Distribution." Ann. Math. Stats., Vol. 9, No. 4, pp. 621-627

Canfield,R.V., Olsen,D.R., Hawkins,R.H. and Chen,T.L., 1980. "Use of Extreme Value Theory in Estimating Flood Peaks from Mixed Populations." Utah Water Research Laboratory, Utah State University.

Chow,V.T., 1964. "Handbook of Applied Hydrology." McGraw-Hill.

Dodd, E. L., 1923. "The Greatest and the Least Variate under General Laws of Error." Trans. Am. Math. Sec 25, pp. 525-539

Fisher, R. A., and Tippett, L. H. C., 1928. "Limiting forms of the Frequency Distribution of the Largest or Smallest Member of a Sample," Proc. Cambridge Phil. Soc., Vol. 24, pt. 2. pp. 180-190

Frechet, M., 1927. "Sur la loi de probabilite de l'ecart maximum," Ann. Soc. Polomaise Math. Cracovie 6, pp. 93-116

Gumbel, E. J., 1958. "Statistics of Extremes." Columbia University Press.

Harter, H. L. and Moore, A. H., 1968. "Maximum Likelihood Estimation of the Parameters of Gamma and Weibull Population from Complete and from Consored Sample." Technometrics. Vol. 7, No. 6, pp. 639-643

Kendall, M.G. and Stuart, A., 1958. "The Advanced Theory of Statistics," Hafner Publishing Comp., N.Y. N.Y.

Kimball, B. F., 1946. "Sufficient Statistical Estimation Functions for the Parameters of the Distribution of Maximum Values." Ann. Math. Stats., vol. 17, pp. 299-309.

Kwan, P.K., 1979. "Weighted Least Squares Estimation of Extreme Value Distribution." M. S. Thesis, Dept. of Applied Statistics and Computer Science, Utah State University, Logan, Utah.

Langbein, W. B., 1949. "Annual Floods and the Partial Duration Flood Series," Trans. Am. Geophys. Union, vol. 30, pp. 879-881

von Mises, R., 1936. "La distribution de la plus grande de n values," Revue Math. l'Union Interbalcanique 1. pp. 1-20

Panchang, G. M., 1967. "Improved Precision of Future High Flood." Proc. Symp. on Floods and their Computations." Leningrad. USSR, pp. 51-59.

Tippett, L. H. C., 1925. "On the Extreme Individuals and the Range of Samples taken from a Normal Population, Biometrika 17, Pts. 3 and 4, pp.364-387

Wallis, J. R., Hosking, J. R. M. and Wood, E. F., 1985. "Estimation of the Generalized Extreme Value Distribution by the Method of Probility Weighted Moments." Technometrics, 27, pp. 251-261.

White, J. S., 1969. "The Moments of Log-Weibull Order Statistics." Technometrics, 11(2): pp.373-385

APPENDIX

```
C%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
C%                                                                       %
C%      THIS IS MAIN PROGRAM WHICH IS USED TO ESTIMATE                    %
C%      THE TWO PARAMETERS OF EXTREME VALUE DISTRIBUTION TYPE I           %
C%      "GUMBEL DISTRIBUTION" USING THE OBSERVED ANNUAL FLOOD             %
C%      DATA.                                                             %
C%                                                                        %
C%      INPUT DATA:                                                       %
C%              X   :     OBSERVED ANNUAL FLOOD                           %
C%              NUMQ:     SAMPLE SIZE                                      %
C%              ITER:     USING THE MAXIMUM LIKELIHOOD METHOD (           %
C%                        NUMERICAL METHOD EVALUATED BY PANCHANG)         %
C%                        ITERATION 4 TIMES IS CLOSED ENOUGH FOR          %
C%                        THE ERROR LIMITATION.                           %
C%                                                                        %
C%      OUTPUT     :                                                      %
C%              A   :     THE SCALE PARAMETER OF THE GUMBEL               %
C%                        DISTRIBUTION.                                   %
C%              B   :     THE SHAPE PARAMETER OF THE GUMBEL               %
C%                        DISTRIBUTION                                    %
C%                                                                        %
C%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
C
 PROGRAM ESTIMA
      REAL*8 PEAKS(1000),A, B
      INTEGER NUMQ,ITER
C
C     INPUT THE OBSERVED ANNUAL FLOOD DATA, SAMPLE SIZE
C     HERE WE USE 59, AND THE NUMBER OF THE ITERATIONS FOR MAXIMUM
C     LIKELIHOOD METHOD (HERE USING 4 ITERATIONS ).
C
      OPEN(10,FILE='LIKE.DAT',STATUS='OLD')
      READ(10,*) NUMQ,ITER
      READ(10,*) ( PEAKS(I),I=1,NUMQ)
C
C     THIS SUBROUTINE MAXL IS USED TO ESTIMATE THE TWO
C     PARAMETERS, SCALE PARAMETER A AND SHAPE PARAMETER B,
C     OF THE GUMBEL DISTRIBUTION USING THE MAXIMUM
C     LIKELIHOOD METHOD WHICH IS NUMERICALIZED BY PANCHANG.
C
      CALL MAXL(PEAKS,NUMQ,ITER,A,B)
      WRITE(6,*) A,B
      STOP
      END
```

```
C%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
C%                                                                  %
C%        THIS IS THE MAIN PROGRAM TO CALL THE SUBROUTINE,          %
C%        RANDOM, MAXL, AND CALCULATE THE AVERAGE DIFFERENCE AND    %
C%        MEAN SQUARE TO TEST THE ACCURACY OF THE ESTIMATION        %
C%        WHICH IS USING THE MONTE CARLO METHOD.                    %
C%                                                                  %
C%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
C%                                                                  %
C%        INPUT DATA:                                               %
C%            X   :     GENERATED ANNUAL FLOOD DATA                 %
C%            ITER:     THE ITERATION FOR MAXIMUM LIKELIHOOD        %
C%                      ESTIMATION FOR THE PARAMETERS OF THE        %
C%                      EXTREME VALUE DISTRIBUTION                  %
C%            A   :     THE SCALE PARAMETER OF THE EXTREME          %
C%                      VALUE DISTRIBUTION ESTIMATED FROM THE       %
C%                      OBSERVATION DATA                            %
C%            B   :     THE SHAPE PARAMETER OF THE EXTREME          %
C%                      VALUE DISTRIBUTION ESTIMATED FROM THE       %
C%                      OBSERVATION DATA                            %
C%                                                                  %
C%        VARIABLES :                                               %
C%            MEAN:     THE MEAN OF THE GENERATED DATA              %
C%            ASSE:     THE AVERAGE OF THE SQUARE OF THE            %
C%                      STANDARD ERROR OF THE PARAMETER A           %
C%            BSSE:     THE AVERAGE OF THE SQUARE OF THE            %
C%                      STANDARD ERROR OF THE PARAMETER B           %
C%            AVA :     AVERAGE OF THE ESTIMATED PARAMETER          %
C%                      A BY MONTE CARLO METHOD                     %
C%            BVB :     AVERAGE OF THE ESTIMATED PARAMETER          %
C%                      B BY MONTE CARLO METHOD                     %
C%            AMSE:     ASSE DIVIDED BY MC(MONTE CARLO NUMBER)      %
C%            BMSE:     BSSE DIVIDED BY MC                          %
C%            MC  :     MONTE CARLO NUMBER (HERE USING 500 )        %
C%            NUMQ:     SAMPLE SIZE (HERE USING 59 )                %
C%            ISEED:    8764321                                     %
C%                                                                  %
C%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
C
        PROGRAM MAIN
        REAL*8 PEAKS(1000), MEAN, ASSE, BSSE, A, B, AVA, BVB
     *         ,AMSE, BMSE, X(1000)
        INTEGER MC,ITER,NUMQ
C
C       INPUT DATA : NUMQ,ITER,A,B
C
        OPEN(10,FILE='DATAIN.DAT',STATUS='OLD')
C
C       OUTPUT :  AVA.BVB,AMSE.BMSE
C
        OPEN(11,FILE='DATA.OUT',STATUS='NEW')
```

```
      READ(10,*) A,B,ISEED,NUMQ,ITER
      WRITE(11,300) NUMQ
      READ(5,*) MC
      WRITE(11,400) MC
      WRITE(11,450)
C
C%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
C%                                                                    %
C%      THE SUBROUTINE RANDOM IS USED TO GENERATE THE                 %
C%      RANDOM NUMBER OF GUMBEL DISTRIBUTION (EXTREME VALUE           %
C%      DISTRIBUTION TYPE I ) WHICH IS SO CALLED THE                  %
C%      ANNUAL FLOOD.                                                 %
C%                                                                    %
C%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
C
      DO 100 I=1,MC
      CALL RANDOM(X,NUMQ,A,B,ISEED)
C
C%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
C%                                                                    %
C%      THE SUBROUTINE MAXL IS USED TO ESTIMATE THE PARAMETERS        %
C%      FOR GUMBEL DISTRIBUTION EA AND EB USING THE DATA WHICH        %
C%      ARE GENERATED BY THE SUBROUTINE RANDOM AS THE ANNUAL          %
C%      FLOOD DATA                                                    %
C%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
C
      CALL MAXL(X,NUMQ,ITER,EA,EB)
      ASSE=0.0
      BSSE=0.0
      AVA=0.0
      BVB=0.0
      ASSE=ASSE+(A-EA)*(A-EA)
      BSSE=BSSE+(B-EB)*(B-EB)
      AVA=AVA+A
      BVB=BVB+B
      WRITE(11,500) EA,EB
  100 CONTINUE
      AMSE=ASSE/MC
      BMSE=BSSE/MC
      AVA=AVA/MC
      BVB=BVB/MC
  300 FORMAT(5X,'SAMPLE SIZE=     ',3X,I4/)
  400 FORMAT(5X,'MONTE CARLO NUMBER=  ',3X,I4/)
  500 FORMAT(5X,E18.8,5X,F12.4)
      WRITE(11,600) AMSE,BMSE,AVA,BVB
  450 FORMAT(9X,'ESTIMATE A',
     *9X,'ESTMATE B'/)
  600 FORMAT(5X,'AMSE=',1X,E10.4,5X,'BMSE=',3X,F8.2,3X,'AVA=',
     *   3X,E10.4,3X,'BVB=',3X,F8.2)
      STOP
      END
```

```
C%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
C                                                                  %
C      THIS IS THE SUBROUTINE WHICH IS USED TO ESTIMATE THE TWO    %
C      PARAMETERS OF THE GUMBEL DISTRIBUTION AK AND U. THE DATA    %
C      ARE GENERATED BY THE RANDOM SUBROUTINE.                     %
C      INPUT:                                                      %
C          X      :   GENERATED ANNUAL FLOOD                       %
C          NUMB   :   SAMPLE SIZE                                  %
C          ITER   :   THE ITERATION NUMBER IS 4.  THE ERROR IS     %
C                     WITHIN THE LIMIT.                            %
C          a , B  :   THE TWO PARAMETERS  ARE  ESTIMATED  FROM     %
C                     THE MAIN PROGRAM EST.                        %
C                                                                  %
C%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
C
       SUBROUTINE MAXL(X,NUMB,ITER,EA,EB)
       INTEGER T(9)
       REAL*8 PEAKS(500), MEAN, Q(9), SE(9), R(9)

       OPEN(11,FILE='LIKE.OUT',STATUS='NEW')

       RNUMQ=NUMQ

    1 FORMAT(12F10.0)
      SUMX=0.0
      DO 11 I=1,NUMQ
   11 SUMX=SUMX+PEAKS(I)
      MEAN=SUMX/RNUMQ
      SUMSD=0.0
      DO 22 I=1,NUMQ
   22 SUMSD=(PEAKS(I)-MEAN)**2+SUMSD
      VAR=SUMSD/(RNUMQ-1)
      SD=SQRT(VAR)
C  COMPUTE FIRST APPROX OF PARAMETER
      AK=3.1416/(DSQRT(6.0)*SD)
      WRITE(11,2) MEAN,VAR,SD,AK
    2  FORMAT('1',4F20.5)
C START ITERATIVE LOOP TO GET U
      KOUNT=0
 2222 S2=AK*0.43429
      S3=1.0/AK
      S4=S3**2
      S5=MEAN-S3
      SUM6=0.0
      SUM7=0.0
      SUM8=0.0
      DO 33 I=1,NUMQ
      DUM6=DEXP(-AK*PEAKS(I))
      SUM6=SUM6+DUM6
      SUM7=SUM7+DUM6*PEAKS(I)
   33 SUM8=SUM8+DUM6*PEAKS(I)**2
```

```
      FAK=SUM7-S5*SUM6
      FPAK=S5*SUM7-SUM8-S4*SUM6
      HK=-FAK/FPAK
      AK=AK+HK
C  PRINT OUT ALL STEP
      WRITE(11,3) AK,S2,S3,S4,S5,SUM6,
     *          SUM7,SUM8,FAK,FPAK,HK
    3 FORMAT(4X,5F14.0/4X,6F16.6)
      KOUNT=KOUNT+1
      IF(KOUNT .GE. ITER) GOTO 9999
      GO TO 2222
 9999 A=S3*DLOG(10.)*DLOG10(RNUMQ)
      B=S3*DLOG(10.)*DLOG10(SUM6)
      UK=A-B
      WRITE(11,4) AK,UK
    4 FORMAT(4X, 'ALPHA= ', F10.7,' U= ',F15.7)
C    COMPUTE Q AND STANDARD ERROR FOR DIFFERENT RETURN PERIOD
      T(1)=2
      T(2)=5
      T(3)=10
      T(4)=20
      T(5)=50
      T(6)=100
      T(7)=200
      T(8)=500
      T(9)=1000
      DUM=S3/SQRT(RNUMQ)
      WRITE(11,5)
    5 FORMAT(4X,'    T=',5X,'   Q=', 4X,'      S.E.='/)
      DO 44 I=1,9
      R(I)=T(I)
      DUMT=ALOG(R(I)/R(I)-1))
      Q(I)=UK-S3*ALOG(DUMT)
      DUMSE=(6.0/3.1416**2)*(1.0-.577216-ALOG(DUMT))
      SE(I)=DUM*(1.0+DUMSE)*1.64
   44 FORMAT(11,6) T(I),Q(I),SE(I)
    6 FORMAT(' ',I5,2F10.0)
      RETURN
      END
```

```
C%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
C%                                                                %
C%      THIS SUBROUTINE RANDOM IS USED TO GENERATE THE            %
C%      RANDOM NUMBER OF GUMBEL DISTRIBUTION (EXTREME VALUE       %
C%      DISTRIBUTION TYPE I) WHICH IS SO CALLED THE ANNUAL        %
C%      FLOOD DATA (GENERATED DATA).                              %
C%                                                                %
C%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
C
      SUBROUTINE RANDOM(X,N,A,B,ISEED)
      REAL*8 X(1000),TEMP,A,B,Y,U
      INTEGER N,I
      TEMP=0.
      DO 10 I=1,N
      Y=RAN(ISEED)
      U=1.0-(1.0-TEMP)*Y**(1.0/(FLOAT(N-I)+1.0))
      X(I)=B-DLOG(-DLOG(U))/A
      TEMP=U
   10 CONTINUE
      RETURN
      END
```