Computer Program Generation of Extreme Value Distribution Data

Stephen (Wan-tsing) Lei
Utah State University

Follow this and additional works at: https://digitalcommons.usu.edu/gradreports

Part of the Applied Statistics Commons

Recommended Citation
Lei, Stephen (Wan-tsing), "Computer Program Generation of Extreme Value Distribution Data" (1986). All Graduate Plan B and other Reports. 1168.
https://digitalcommons.usu.edu/gradreports/1168

This Report is brought to you for free and open access by the Graduate Studies at DigitalCommons@USU. It has been accepted for inclusion in All Graduate Plan B and other Reports by an authorized administrator of DigitalCommons@USU. For more information, please contact digitalcommons@usu.edu.
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
ACKNOWLEDGEMENTS

I would like to express sincere appreciation to my major professor, Dr. Ronald V. Canfield for the many hours of precious help he extended to me during my studies. As a friend and teacher he made this time at Utah State University both enjoyable and educational.

Thanks also to my committee members Dr. Kenneth S. Lyon and Dr. David L. Turner for their advice during my graduate study.

I would also like to thank Dr. Rex L. Hurst, Dr. Donald V. Sisson, and Dr. David White who made my education at Utah State University a pleasant and successful one.

Special thanks to my wife, Lung Shiang (Mary), who contributed to my education as a living earner, mother of three children, housewife, and frequent evening widow, and gave me the support and encouragement needed to survive and complete my masters program, my love and affection.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGMENTS</td>
<td>ii</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>iv</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>v</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>vi</td>
</tr>
<tr>
<td>I. INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>Relevance of Investigation</td>
<td>1</td>
</tr>
<tr>
<td>Objective of Study</td>
<td>3</td>
</tr>
<tr>
<td>II. LITERATURE REVIEW</td>
<td>5</td>
</tr>
<tr>
<td>Data Sampling</td>
<td>5</td>
</tr>
<tr>
<td>Extreme Value Theory</td>
<td>5</td>
</tr>
<tr>
<td>Parameters Estimation</td>
<td>7</td>
</tr>
<tr>
<td>Monte Carlo Method</td>
<td>11</td>
</tr>
<tr>
<td>III. METHODOLOGY</td>
<td>12</td>
</tr>
<tr>
<td>Maximum Likelihood Method</td>
<td>12</td>
</tr>
<tr>
<td>Monte Carlo Experiment</td>
<td>14</td>
</tr>
<tr>
<td>Model Accuracy Test</td>
<td>17</td>
</tr>
<tr>
<td>IV. EXAMPLE</td>
<td>19</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>21</td>
</tr>
<tr>
<td>APPENDIX</td>
<td>23</td>
</tr>
<tr>
<td>Table</td>
<td>Description</td>
</tr>
<tr>
<td>-------</td>
<td>-------------</td>
</tr>
<tr>
<td>1.</td>
<td>Observed Data for Estimating the Gumbel Parameters</td>
</tr>
<tr>
<td>2.</td>
<td>Using Mean Squared Error to Measure the Accuracy of the Monte Carlo Method</td>
</tr>
</tbody>
</table>
# LIST OF FIGURE

<table>
<thead>
<tr>
<th>Figures</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. The Flow Chart of the Program</td>
<td>18</td>
</tr>
<tr>
<td>2. Schematic Showing Data Collection</td>
<td>19</td>
</tr>
</tbody>
</table>
ABSTRACT

COMPUTER PROGRAM 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 analysis of complex problems which cannot be solved analytically. The Gumbel distribution is an important distribution in hydrology. It is used to describe 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. Analysis 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.
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 occurrence 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 distributions 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:

\[
P(X) = P(x_v \leq y)
\]

(2-1)

2. Define \( X_N \) as the maximum value of \( x \) in a sample of
length \( N \), i.e., \( X_N = \max_{1 \leq v \leq N} x_v \) so that

\[
P(X_N \leq y) = P(x_1 \leq y, x_2 \leq y, \ldots, x_N \leq y)
\]
or \( P(X_N \leq y) = [P(y)]^N \tag{2-2} \)

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

\[
P(y) = 1 - ae^{-y} \tag{2-3}
\]

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

\[
P(X_N \leq y+\ln(aN))=[P(y+\ln(aN))]^N \tag{2-4}
\]
and from Equation (2-3)

\[
P(y + \ln(aN))=1-ae^{-(y + \ln(aN))} \tag{2-5}
\]
so that

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

\[
P(X_N \leq y+\ln(aN))=[1 - e^{-Y/N}]^N \tag{2-6}
\]

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

\[
\lim_{N \to \infty} P(x_N \leq y+\ln(an))=\exp(-e^{-Y}) \tag{2-7}
\]

\[
P(x) =\exp(-e^{-Y})
\]
or

\[
P(x) = \exp(-e^{-a(x + B)}) \tag{2-8}
\]

**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:

\[ U'_y, r = \int_{-\infty}^{\infty} y^r \exp(-y-e)^-Y dy \quad (2-9) \]

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

\[ U'_y, r = \int_{-\infty}^{\infty} (-\ln z)^r e^{-z} (-dz/z) \quad (2-10) \]

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

\[ U'_y, 1 = \int_0^{\infty} \ln z(z) e^{-z} dz/z \quad (2-11) \]

But \( \int_0^{\infty} ze^{-z} dz \) is \( \Gamma(1) \), so that, from Kendall and Stuart (1958)

\[ U'_y, 1 = -\psi(1) = \psi \]

where \( \psi \) is Euler's constant, approximately 0.57721.

Reconverting to the original variate \( x \) as \( x = y/a + B \)

\[ U'_1 = B + \psi/a \quad (2-12) \]

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

\[ u_2 = \frac{\pi^2}{6a^2} \quad \sigma = \frac{\pi}{\sqrt{6}a} \]

\[ a = \frac{\pi}{\sqrt{6}\sigma} = 1.2825 /\sigma \quad (2-13) \]

\[ B = u - \frac{\psi}{a} = u - (0.57721/a) \quad (2-14) \]

"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,

\[ P(x) = \exp(-e^{-y}) \]  \hspace{1cm} (2-15)

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

\[ a \ (X_{in} - B) = \log (-\log P(x)) = y_i \]  \hspace{1cm} (2-16)

\[ y_i = f(a, B, X_{in}) \]  \hspace{1cm} (2-17)

\[ z = \sum (E_i - y_i)^2 w_i \]  \hspace{1cm} (2-18)

Thus

\[ z = \sum (a \ X_{in} - a \ B - E_i)^2 w_i \]  \hspace{1cm} (2-19)

\[ Q(X_i) = X_i \]  \hspace{1cm} (2-20)

where

- \( E_i \) is known value, \( E_i = E(Y_i) \)
- \( W_i \) is the weight function of \( Y_i \), \( W_i = [\text{Var}(Y_i)]^{-1} \)
- \( X_{in} \) 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,

\[ Z = \sum [h_1(a, B) + h_2(a, B)Q(X_{in}) - E_i]^2 w_i \]  \hspace{1cm} (2-21)

\[ h_2(a, B) = \frac{\sum (Q(X_{in}) - \bar{Q})(E_i - \bar{E}) W_i}{\sum (Q(X_{in}) - \bar{Q})^2 W_i} \]  \hspace{1cm} (2-22)
\[ h_1(a, B) = \bar{E} = h_2(a, B) \bar{Q} \]  
(2-23)

\[ \bar{E} = \frac{\sum E_i W_i}{\sum W_i} \]  
(2-24)

\[ \bar{Q} = \frac{\sum Q(X_{in}) W_i}{\sum W_i} \]  
(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 $y_1, ..., y_N$ is defined as the joint probability density $f(y_1, ..., y_N; Q)$. The likelihood $L(Q)$ is viewed as a function, an arbitrary value of the distribution parameter. The true value (unknown) is denoted by $Q_0$. 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,

$$L(Q) = f(y_1; Q) f(y_2; Q) ... f(y_n; Q) \quad (3-1)$$
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,

\[
f(X_1) = ae^{-a(X_1-B)}\exp(-e^{-a(X_1-B)})
\]

same as for \( X_2 \)

\[
f(X_2) = ae^{-a(X_2-B)}\exp(-e^{-a(X_2-B)})
\]

In general terms

\[
f(X_1, \ldots X_N) = a^N e^{-\sum a(X_i-B)}\exp(-e^{-\sum a(X_i-B)})
\]

\[
L = \log f(X_1, \ldots X_N) = \sum \log a(X_i-B) - \sum e^{-a(X_i-B)}
\]

Then maximize the likelihood function as follows;

\[
\frac{\partial L}{\partial a} = 0 \quad \text{and} \quad \frac{\partial L}{\partial B} = 0
\]

\[
B = (\log_{10} \log_{10} N) / a - (\log_{10} \log_{10} (\Sigma e^{-aX_i})) / a
\]

\[
F(a) = \Sigma X_i e^{-aX_i} - (\bar{X} - 1/a) \Sigma e^{-aX_i} = 0
\]

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

\[
\frac{dF(a)}{d(a)}
\]

\[
F'(a) = -\Sigma X_i^2 e^{-aX_i} + (\bar{X} - 1/a) \Sigma X_i e^{-aX_i} - (1/a)^2 \Sigma e^{-aX_i}
\]

Panchang (1967) used the successive approximations to estimate
a_1 by applying the following equation.

\[
\frac{-F(a_1)}{F'(a_1)} = h_1
\]

(3-10)

then, \(F(a_2) = F(a_1 + h_1) = F(a_1) + h_1 F'(a_1)\) \hspace{1cm} (3-11)

and \[a_2 = a_1 + h_1\] \hspace{1cm} (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 pseudo-random numbers was generated (equivalent to a single hypothetical flood record for a site).

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

3. From a known distribution (Gumbel distribution) 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 referred 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). $\ln[-\ln F(X)] = -a(X-B) = \ln[-\ln(U)]$ (3-15)
d). $X = -\ln[-\ln(U)]/a + B$ (3-16)

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

$$X(I) = B - \text{DLOG}[-\text{DLOG}(U)]/a$$

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

$$U = 1.0 - (1.0-\text{TEMP})*\text{Y}^{(1.0/(\text{FLOAT(N-I)}+1.0))}$$

**Model Accuracy Test**

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

$$\text{Error} = \frac{\sum_{\text{observed Sample}} \left( \text{Esti. A by using generating Sample} - \text{Esti. A by using observed Sample} \right)^2}{\text{Number of Monte Carlo Trials}}$$ (3-17)
Sum of Square Error = \frac{\sum [(\text{Esti. B by using generating Sample} - \text{Esti. B by using observed Sample})]^2}{\text{Number of Monte Carlo Trials}} \quad (3-18)

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

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

The flow chart of the Monte Carlo program is shown on Figure 1.
INPUT: NUMQ=SAMPLE SIZE  
ITER=NUMBER OF ITERATION  
OBSERVED ANNUAL PEAKS:  
PEAKS(I), I = 1, NUMQ

KOUNT<ITER  

ESTIMATE PARAMETERS A AND B

OUTPUT: ESTIMATED PARAMETERS A AND B

INPUT: A, B, ITER,  
NUMQ, AND MC (MONTE CARLO REPEATEITIONS)

I < MC  

GENERATING RANDOM VARIATE AND ESTIMATING EA AND EB

OUTPUT:  
AMSE = ASSE/MC  
BMSE = BSSE/MC  
AVA = AVA /MC  
BVB = BVB /MC

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

<table>
<thead>
<tr>
<th>Year</th>
<th>Annual floods</th>
<th>Streamflow</th>
<th>X1</th>
<th>X2</th>
<th>X3</th>
<th>X4</th>
<th>Xn-2</th>
<th>Xn</th>
</tr>
</thead>
<tbody>
<tr>
<td>65</td>
<td></td>
<td></td>
<td>10000</td>
<td>10700</td>
<td>20100</td>
<td>8210</td>
<td>14300</td>
<td>8040</td>
</tr>
<tr>
<td>1390</td>
<td></td>
<td></td>
<td>18500</td>
<td>13000</td>
<td>16400</td>
<td>14500</td>
<td>13000</td>
<td>17200</td>
</tr>
<tr>
<td>11900</td>
<td></td>
<td></td>
<td>13600</td>
<td>12400</td>
<td>18300</td>
<td>12900</td>
<td>18200</td>
<td>9900</td>
</tr>
<tr>
<td>9020</td>
<td></td>
<td></td>
<td>11800</td>
<td>16100</td>
<td>16900</td>
<td>11800</td>
<td>13900</td>
<td>12300</td>
</tr>
<tr>
<td>11900</td>
<td></td>
<td></td>
<td>11900</td>
<td>16000</td>
<td>11500</td>
<td>19900</td>
<td>18600</td>
<td>18000</td>
</tr>
<tr>
<td>29100</td>
<td></td>
<td></td>
<td>10300</td>
<td>12200</td>
<td>15600</td>
<td>12700</td>
<td>13100</td>
<td>19200</td>
</tr>
<tr>
<td>23000</td>
<td></td>
<td></td>
<td>8700</td>
<td>7150</td>
<td>14300</td>
<td>20600</td>
<td>25600</td>
<td>8180</td>
</tr>
<tr>
<td>16100</td>
<td></td>
<td></td>
<td>10200</td>
<td>12300</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Results: Using the observed data,

Estimator \( a = 0.2375 \times 10^{-3} \)

Estimator \( B = 12156.98 \)

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        | \( .237 \times 10^{-3} \)  | \( .237 \times 10^{-3} \)  | \( .237 \times 10^{-3} \)  | \( .237 \times 10^{-3} \)  |
| Estimate of Parameter EB        | \( 12561.98 \)   | \( 12400.12 \)   | \( 12278.55 \)   | \( 12181.29 \)   |
| Mean Squared Error of a, slope para. | \( .668 \times 10^{-12} \)  | \( .1098 \times 10^{-12} \)  | \( .495 \times 10^{-12} \)  | \( .886 \times 10^{-12} \)  |
| Mean square error of B, location para. | \( 385.01 \)  | \( 4707.85 \)  | \( 6.07 \)  | \( 75.80 \)  |
| Bias in Estimator a, shape para. | \( .791 \times 10^{-5} \)  | \( .475 \times 10^{-5} \)  | \( .237 \times 10^{-5} \)  | \( .475 \times 10^{-6} \)  |
| Bias in Estimator B, location para. | \( 405.23 \)  | \( 243.14 \)  | \( 121.57 \)  | \( 24.31 \)  |
REFERENCES


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


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% INPUT DATA:
C% X : OBSERVED ANNUAL FLOOD
C% NUMQ: SAMPLE SIZE
C% ITER: USING THE MAXIMUM LIKELIHOOD METHOD (NUMERICAL METHOD EVALUATED BY PANCHANG)
C% ITERATION 4 TIMES IS CLOSED ENOUGH FOR THE ERROR LIMITATION.

C% OUTPUT :
C% A : THE SCALE PARAMETER OF THE GUMBEL DISTRIBUTION.
C% B : THE SHAPE PARAMETER OF THE GUMBEL DISTRIBUTION.

C% PROGRAM ESTIMA
REAL*8 PEAKS(1000),A, B
INTEGER NUMQ, ITER

C INPUT THE OBSERVED ANNUAL FLOOD DATA, SAMPLE SIZE HERE WE USE 59, AND THE NUMBER OF THE ITERATIONS FOR MAXIMUM LIKELIHOOD METHOD (HERE USING 4 ITERATIONS).

OPEN(10,FILE='LIKE.DAT',STATUS='OLD')
READ(10,*) NUMQ, ITER
READ(10,*) (PEAKS(I), I=1, NUMQ)

C THIS SUBROUTINE MAXL IS USED TO ESTIMATE THE TWO PARAMETERS, SCALE PARAMETER A AND SHAPE PARAMETER B, OF THE GUMBEL DISTRIBUTION USING THE MAXIMUM LIKELIHOOD METHOD WHICH IS NUMERICALIZED BY PANCHANG.

CALL MAXL(PEAKS, NUMQ, ITER, A, B)
WRITE(6,*) A, B
STOP
END
C% THIS IS THE MAIN PROGRAM TO CALL THE SUBROUTINE, RANDOM, MAXL, AND CALCULATE THE AVERAGE DIFFERENCE AND MEAN SQUARE TO TEST THE ACCURACY OF THE ESTIMATION WHICH IS USING THE MONTE CARLO METHOD.

INPUT DATA:
X : GENERATED ANNUAL FLOOD DATA
ITER: THE ITERATION FOR MAXIMUM LIKELIHOOD ESTIMATION FOR THE PARAMETERS OF THE EXTREME VALUE DISTRIBUTION
A : THE SCALE PARAMETER OF THE EXTREME VALUE DISTRIBUTION ESTIMATED FROM THE OBSERVATION DATA
B : THE SHAPE PARAMETER OF THE EXTREME VALUE DISTRIBUTION ESTIMATED FROM THE OBSERVATION DATA

VARIABLES:
MEAN: THE MEAN OF THE GENERATED DATA
ASSE: THE AVERAGE OF THE SQUARE OF THE STANDARD ERROR OF THE PARAMETER A
AVA : AVERAGE OF THE ESTIMATED PARAMETER A BY MONTE CARLO METHOD
BVB : AVERAGE OF THE ESTIMATED PARAMETER B BY MONTE CARLO METHOD
AMSE: ASSE DIVIDED BY MC(MONTE CARLO NUMBER)
BMSE: BSSE DIVIDED BY MC
MC : MONTE CARLO NUMBER (HERE USING 500)
NUMQ: SAMPLE SIZE (HERE USING 59)
ISEED: 8764321

PROGRAM MAIN
REAL*8 PEAKS(1000), MEAN, ASSE, BSSE, A, B, AVA, BVB, AMSE, BMSE, X(1000)
INTEGER MC,ITER,NUMQ

INPUT DATA : NUMQ,ITER,A,B

OPEN(10,FILE='DATAIN.DAT',STATUS='OLD')
OUTPUT : AVA,BVB,AMSE,BMSE

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% 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
DO 100 I=1,MC
CALL RANDOM(X,NUMQ,A,B,ISEED)

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,'ESTIMATE 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
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
FPACK=S5*SUM7-SUM8-S4*SUM6
HK=-FAK/FPACK
AK=AK+HK

C PRINT OUT ALL STEP
WRITE(11,3) AK,S2,S3,S4,S5,SUM6,
* SUM7,SUM8,FAK,FPACK,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=',SX, 1 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
THIS SUBROUTINE RANDOM IS USED TO GENERATE THE RANDOM NUMBER OF GUMBEL DISTRIBUTION (EXTREME VALUE DISTRIBUTION TYPE I) WHICH IS SO CALLED THE ANNUAL FLOOD DATA (GENERATED DATA).

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