VTFIT

A Routine for Fitting Homogeneous
Probability Density Functions

User Documentation

Richard Cooke
Department of Agricultural Engineering
Virginia Polytechnic Institute
Blacksburg, VA 24061-0303

July, 1993

INTRODUCTION

Introducing VTFIT

VTFIT was developed to fit probability distribution functions to data. The routine provides maximum likelihood estimates of the parameters of the

- Gaussian (Normal)
- Two-Parameter Log Gaussian (Log Normal)
- Three-Parameter Log Gaussian
- Gamma
- Three-Parameter Gamma (Pearson Type III)
- Log Gamma (Log Pearson Type III)
- Inverted Gamma (Pearson Type V)
- Gumbel (Extreme Value Type I) For Minima
- Gumbel (Extreme Value Type I) For Maxima
- Frechet (Extreme Value Type II) For Minima
- Frechet (Extreme Value Type II) For Maxima
- Three Parameter Frechet For Maxima
- Weibull (Extreme Value Type III) For Maxima
- Weibull (Extreme Value Type III) For Minima
- Three-Parameter Weibull For Minima
- Exponential
- Shifted Exponential, and the
- Beta

distributions.

Test statistics are provided for the

- Chi-Square
- Kolmogorov-Smirnov
- Kuiper
- Cramer-Von Mises
- Anderson-Darling, and the
- Maximum Likelihood

goodness-of-fit tests.

Graphs of the frequency histogram with the fitted probability distribution function superimposed, and of the empirical distribution function for the sample data, together with the fitted cumulative density function, are provided for visual assessment.

Intended Users

This manual is intended for users who have, in addition to a knowledge of distribution fitting theory, a rudimentary knowledge of the disk operating system (MS-DOS or PC-DOS) as implemented on IBM and IBM compatible computers. Users should be able to

- Create Data Files
- Edit Data Files
- Create Subdirectories, and
- Transfers Files Between Subdirectories

System Requirements

This minimum system requirements include and IBM or IBM compatible computer with a graphics adapter and 256k of free memory. Additional recommended requirements are a hard disk with as least 1 MB of free memory for program installation, an 80286 or faster processor, an 80x87 math co-processor, and a color monitor.

Keyboard Conventions

In this document <enter> represents the key labeled "<┘ Enter" , <up arrow> represents the key labeled "↑", and <down arrow> represents the key labeled "↓" on the computer keyboard.

Acknowledgments

This routine incorporates procedures from the WINDOW TOOLS routines developed by J. Morgan and the STATTAB routines developed by B. Brown. Both sets of routines have been placed by their authors in the public domain.

Disclaimer

No warranty of any kind, either expressed or implied, is provided. All decisions as to the suitability of VTFIT to any task are left up to the user. The user is responsible for any and all costs associated with the use of VTFIT, including incidental and/or consequential costs accruing from this use.

DISTRIBUTION FITTING THEORY

Choosing a distribution

The usual approach to distribution fitting is to fit as many distributions as possible and use goodness-of-fit tests to determine the best fit. This method, the empirical method, is subjective and is not always conclusive. However, except in the case of data which represent the extrema of processes, there is no single accepted rule for selecting one distribution over another.

For extreme data, the choice of a distribution may be narrowed down to choosing one of three possible distributions. These three distributions are generically known as the extreme value type I, type II and type III distributions, but they are also referred to as the Gumbel, Frechet and Weibull distributions, respectively (Castillo, 1988). Haan (1977) and Canfield et al. (1981) have also referred to the extreme value type II as the Cauchy type extreme value distribution. Each of the extreme value distributions may be written in two forms--- one bounded on the left and the other on the right. One form is suitable for data representing the maximum values of a process and the other for minimum values. Each form can be transformed into the other by replacing the fitted variable, *x*, by -*x* and suitably adjusting the location parameter.

Lowing (1987) recommended a procedure for selecting a probability distribution function for hydrologic data which need not be extreme-value data. In this procedure, information relating to the physical nature of the variable being fitted and the skew of the sample set are used to narrow down the choice of a distribution, then goodness-of-fit tests are used to select the best fitting distribution from a reduced set of distributions.

Estimating distribution parameters

The preferred method of parameter estimation for distribution fitting is the maximum likelihood method (Law and Kelton, 1991). This method consists of maximizing the likelihood function, *L*, given by

$$\mathcal{L} = \prod_{i=1}^{n} f(x_i) \qquad (1)$$

where *f(x)* is the probability density function of the selected distribution, and x_1, x_2, ... x_n are the n data points in the sample to be fitted. More commonly, the log-likelihood function given by

$$\ln(L) = \sum_{i=1}^{n} \ln(f(x_i)) \qquad (2)$$

is maximized. Since the log function is a strictly increasing monotonic function, the results from maximizing either function are identical. For some distributions such as the Gaussian, log-Gaussian and exponential distributions, the maximum likelihood estimates are simple functions of the sample moments (Law and Kelton, 1991; Haan, 1977). For other distributions, the log likelihood function has to be maximized directly or differentiated with respect to the parameters and the resulting equations solved simultaneously. Under certain conditions, the likelihood function may not be properly behaved. The behavior of some distributions in the lower tails, or, in the case of the beta distribution, in both tails, is dependent on the value of the shape parameter. This is clearly demonstrated in the graphical examples of the beta, gamma and Weibull for minima distributions given by Law and Kelton (1991). For shape parameters less than one, these distribution functions are asymptotic to the frequency axis in their lower tails. Thus the likelihood function can be unstable when the shape parameter is close to one and can become infinite for extremely small data values for shape parameters less than one. Castillo (1988) has also pointed out that there are regularity problems with the Weibull function for shape parameters ranging from one to two. Fortunately, maximum likelihood estimates can also be approximated by maximizing the log of the multinomial distribution function which is given by

$$\ln(PMF) = \sum_{i=1}^{k} n_i \ln(P_i) \text{ , and} \qquad (3)$$

$$P_i = \int_{y_i}^{y_{i+1}} f(x)dx; \qquad y_1 < y_2 < \ldots\ldots < y_{k+1} \qquad (4)$$

where y_1, y_2, ... y_{k+1} are k+1 numbers chosen such that all the data points in the sample lie between y_1 and y_{k+1}, and n_i is the number of data points with values ranging from y_i to just less than y_{i+1}. The maximization of this function produces approximate maximum likelihood estimates of the parameters (Castillo, 1988). Although the results obtained are only approximate, the multinomial distribution function has the advantage of being always finite, and the results are quite similar to those obtained from the standard maximum likelihood method (Castillo, 1988).

Evaluating goodness-of-fit

Shapiro and Brain (1981) classified goodness-of-fit tests into three major groups, namely, regression type tests, probability transformation tests and special features tests. The most common regression test is probability plotting, in which the ordered sample data is plotted on a graph in which the axes are transformed so that if the data are in fact conformed to the selected distribution, they lie on a straight line. The deviation of the data from linearity may be visually assessed. Shapiro and Brain (1981) strongly recommended that other goodness-of-fit tests should always be augmented by a probability plot.

An informal visual test described by Law and Kelton (1991) and Woeste et al. (1979) requires that the relative frequencies of the sample data and the fitted probability density function be plotted on the same graph and visually compared. This visual test is a variation of the probability plotting test. The chi-square test, the oldest of all goodness-of-fit tests (Shapiro and Brain, 1981; Law and Kelton, 1991) is a less subjective comparison of frequency histograms with fitted distributions. In this procedure, the range of the sample data is divided into a discrete number of intervals and the number of data points falling in each interval is compared with the expected number predicted by the fitted distribution. The expected number is obtained by integrating the fitted probability distribution between the interval boundaries and multiplying by the number of data points in the sample. Details of this test are given by Law and Kelton (1991) while Chandra (1981) has discussed the effects of correlation, between the data points in the sample, on the test statistic. Although it is less subjective than visual assessment, the chi-square test is not entirely objective. The test statistic is dependent on the number and lengths of the intervals. There is no single accepted rule for choosing either. Law and Kelton (1991) recommended the use of equiprobable intervals with the expected number in each interval being five or more.

Probability transformation tests are based on the fact that if a set of data conforms to a probability distribution, *f(x)*, then the transformed variable, y_i, given by

$$y_i = \int_{-\infty}^{x_i} f(x)dx \qquad (5)$$

conforms to a uniform distribution (Shapiro and Brain, 1981). The test statistic of these tests, which include the Kolmogorov-Smirnov, Kuiper, Cramer-von Mises and the Anderson-Darling tests among others, are measures of the sample deviation from uniformity. In the Kolmogorov-Smirnov procedure, the test statistic is the maximum deviation of the ordered transformed variable either above or below the uniform line, while for the Kuiper procedure the test statistic is the sum of the maximum deviations both above and below the uniform line. The Cramer-von Mises test statistic is essentially the sum of the squared deviations from the uniform line(Shapiro and Brain, 1981), while the Anderson-Darling statistic is a weighted sum of deviations, with more weight given to observations in the tails of the distribution (Law and Kelton, 1991).

Worley et al. (1990) used the numerical value of the log likelihood function as yet another test of goodness-of-fit. They proposed that this statistic be used when other goodness-of-fit procedures fail to discriminate between two or more distributions.

Applying the theory to VTFIT

In VTFIT, the maximum likelihood estimates of the distribution parameters are obtained from the sample moments for those distributions where these estimates are simple functions of the sample moments. For the others, the parameters are estimated by direct maximization of the log likelihood function given in Equation 2 , using an optimization algorithm developed by Rosenbrock and Storey (1966). If, after a fixed number of user specified iterations, the maximization procedure fails to converge, then the parameters are obtained by maximizing the log multinomial distribution function given by Equations 3 and 4. The parameters of the log Pearson type III are evaluated using the method of mixed moments developed by Rao(1980). This fitting procedure has been shown to be superior to the maximum likelihood procedure for this distribution (Arora and Singh, 1989). For those distributions in which the location parameter is estimated, the method presented by Kline and Bender (1990) is used to obtain an initial estimate of the location parameter. The forms of the distributions that can be fitted with VTFIT, and the methods used to evaluate the parameters of these distributions are given in Table 1. The formulae for evaluating the test statistics of the goodness-of-fit tests used in VTFIT are given in Table 2.

Table 1. Distributions in VTFIT

Distribution	Density Function	Parameter Evaluation Method
	Extreme value distributions	
Gumbel (maxima)	$f(x) = \frac{e^{\frac{-(x-\gamma)}{\beta}} e^{-e^{\frac{-(x-\gamma)}{\beta}}}}{\beta}$	B
Gumbel (minima)	$f(x) = \frac{e^{\frac{-(\gamma-x)}{\beta}} e^{-e^{\frac{-(\gamma-x)}{\beta}}}}{\beta}$	B
Frechet (maxima, 2P)	$f(x) = \frac{\alpha}{\beta}\left(\frac{\beta}{x}\right)^{\alpha+1} e^{-\left(\frac{\beta}{x}\right)^{\alpha}}$	A
Frechet (maxima, 3P)	$f(x) = \frac{\alpha}{\beta}\left(\frac{\beta}{x-\gamma}\right)^{\alpha+1} e^{-\left(\frac{\beta}{x-\gamma}\right)^{\alpha}}$	B,C
Frechet (minima)	$f(x) = \frac{\alpha}{\beta}\left(\frac{\beta}{\gamma-x}\right)^{\alpha+1} e^{-\left(\frac{\beta}{\gamma-x}\right)^{\alpha}}$	B,C
Weibull (maxima)	$f(x) = \frac{\alpha}{\beta}\left(\frac{\gamma-x}{\beta}\right)^{\alpha-1} e^{-\left(\frac{\gamma-x}{\beta}\right)^{\alpha}}$	B,C
Weibull (minima, 2P)	$f(x) = \frac{\alpha}{\beta}\left(\frac{x}{\beta}\right)^{\alpha-1} e^{-\left(\frac{x}{\beta}\right)^{\alpha}}$	E
Weibull (minima, 3P)	$f(x) = \frac{\alpha}{\beta}\left(\frac{x-\gamma}{\beta}\right)^{\alpha-1} e^{-\left(\frac{x-\gamma}{\beta}\right)^{\alpha}}$	B,C

Table 1 continued.

Other Distributions		
Gaussian	$f(x)=\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	A
Log-Gaussian (2P)	$f(x)=\frac{1}{x\sqrt{2\pi\sigma^2}}e^{-\frac{(\ln x-\mu)^2}{2\sigma^2}}$	A
Log-Gaussian (3P)	$f(x)=\frac{1}{(x-\gamma)\sqrt{2\pi\sigma^2}}e^{-\frac{(\ln(x-\gamma)-\mu)^2}{2\sigma^2}}$	B, C
Exponential (1P)	$f(x)=\beta e^{-\beta x}$	A
Exponential (2P)	$f(x)=\beta e^{-\beta(x-\gamma)}$	B, C
Beta	$f(x)=\frac{x^{\alpha_1-1}(1-x)^{\alpha_2-1}}{\text{Beta}(\alpha_1,\alpha_2)}$	B
Gamma (2P)	$f(x)=\frac{\beta^{-\alpha}x^{\alpha-1}e^{-\frac{x}{\beta}}}{\Gamma(\alpha)}$	D
Gamma (3P)	$f(x)=\frac{\beta^{-\alpha}(x-\gamma)^{\alpha-1}e^{-\frac{(x-\gamma)}{\beta}}}{\Gamma(\alpha)}$	B, C
Log-Pearson Type III	$f(x)=\frac{e^{\frac{\gamma}{\beta}}x^{\beta-1}}{\lvert\beta\rvert\Gamma(\alpha)}\left(\frac{\ln x-\gamma}{\beta}\right)^{\alpha-1}$	D
Inverted Gamma	$f(x)=\frac{x^{-(\alpha+1)}e^{-\frac{\beta}{x}}}{\beta^{-\alpha}\Gamma(\alpha)}$	D

* **Parameter Notation**

- α: Shape parameter
- β: Scale parameter
- γ: Location parameter

** **Parameter Evaluation Method**

- A: Parameters estimated from sample moments.
- B: Parameters estimated by direct optimization of log-likelihood function.
- C: Initial estimate of location parameter obtained following Kline and Bender (1990).
- D: Parameters estimated from sample moments, then adjusted for moment bais.
- E : Parameters evaluated by solving equations given by Law and Kelton(1991).

Table 2. Goodness-of-fit tests used in VTFIT

Test	Test Statistic
Log-Likelihood	$\ln(L) = \sum_{i=1}^{n} \ln(\hat{f}(x_i))$
Kolmogorov-Smirnov	$D = \max\{D^+, D^-\}$
Kuiper	$V = D^+ + D^-$
Cramer-von Mises	$W^2 = \sum_{i-1}^{n} \left(\hat{F}(x_i) - \frac{2i-1}{2n}\right)^2 + \frac{1}{12n}$
Anderson-Darling	$A^2 = -\frac{1}{n} \sum_{i=1}^{n} \left\{(2i-1)\ln[\hat{F}(X_{(i)})\{1-\hat{F}(X_{(n-i+1)})\}]\right\} - n$
Chi-Square*	$\chi^2 = \sum_{j=1}^{k} \frac{(N_j - np_j)^2}{np_j}$

* For this test the entire range of the fitted distribution has to be first divided into *k* adjacent intervals $[a_0,a_1)$, $[a_1,a_2)$,...,$[a_{k-1},a_k)$

Notation:

$\hat{f}(\ )$:	Fitted probability density function
$\hat{F}(\ )$:	Fitted cumulative distribution function
D^+ :	$\max\left\{\frac{i}{n} - \hat{F}(X_{(i)})\right\}$ for *i* = 1 to *n*
D^- :	$\max\left\{\hat{F}(X_{(i)}) - \frac{i-1}{n}\right\}$ for *i* = 1 to *n*
$X_{(i)}$:	x_i sorted in ascending order
N_i :	Number of x_i's in the *j*th interval $[a_{j-1}, a_j)$
p_j :	$\int_{a_{j-1}}^{a_j} \hat{f}(x)dx$

VTFIT FILES

Input files

A sample input data file is supplied in Figure 1. The first field in each row is an optional index variable. The next three fields are observations from three distinct variables. When the routine is in operation, the user is asked to specify the number of variables in the file and the variable to be fitted, and to indicate if an index variable is present. If present, the index variable should not be considered when specifying the number of variables in the file. If there is more than one variable in a file, and all the variables do not have the same number of observations, then in any row where there is no observation of a variable the quantity "-9999" is inserted in the file. Such values are ignored in the fitting procedure. If the actual value of one observation of a variable is -9999, then this observation should be replaced by a slightly larger or smaller value.

Output files

The output file produced by VTFIT contains the parameters for the fitted distribution and test statistics for the six goodness-of-fit tests outlined in Table 2. It also contains values of observed frequency, the fitted density function, the empirical distribution function, and the fitted cumulative distribution function, for values of the independent variable ranging from the minimum to the maximum observed values. These can be exported to a plotting routine. A sample output file is given in Figure 2.

Figure 1. Sample VTFIT input file

1	1.01	0.92	10.65
2	1.41	0.42	24.45
3	1.11	0.37	6.05
4	0.21	0.74	5.95
5	0.21	0.32	2.9
6	1.61	1.1	11.4
7	1.3	0.31	16.75
8	4.39	0.41	9.7
9	1.26	1.43	7.75
10	1.63	2.1	4.2
11	1.3	1.93	2.6
12	3.01	3.34	0.5
13	0.74	1.9	2.65
14	0.36	1.56	2.65
15	1.16	1.74	6.1
16	1.61	1.19	5
17	0.42	1.37	2.1
18	2.3	0.89	0.55
19	1.41	1.68	8.2
20	2.99	1.15	10.3
21	1.31	0.73	9.85
22	1.53	0.21	12.1
23	2.06	1.16	2.35
24	1	1.73	7
25	1.11	2.9	8.65
26	-9999	0.47	31.9
27	-9999	-9999	2.9
28	-9999	-9999	10.9

Column 1: Index variable (optional)

Columns 2-4: Observations of three distinct variables with one variable per column.

"-9999" added if there is more than one variable and the number of observations of all the variables are not equal

```
****************************************************************************
file.prn  Variable  2  of  3

  Weibull (Extreme value type III) for minima for  26  observations

  Data minimum : .21        Probability of exceedance : .9555005
  Data maximum : 3.34       Probability of exceedance : 1.433009E-02

  Scale Parameter : 1.382663              K-S  D  :  .1124576
  Shape Parameter : 1.639344              Kuiper  :  .186471

  Chi square:  1.692308     DF:  4        p Value:  .7921144
  Log Likelihood: -27.33461               Cramer- von Mises: 3.045515E-02
  Anderson-Darling:  .2844868
```

Variable	Observed Frequency	Fitted PDF	Observed EDF	Fitted CDF
0.21000	0.00000	0.33953	0.00000	0.04450
0.36650	0.73728	0.45292	0.11538	0.10722
0.52300	0.98304	0.51973	0.26923	0.18386
0.67950	0.00000	0.55104	0.26923	0.26805
0.83600	0.49152	0.55449	0.34615	0.35488
0.99250	0.49152	0.53666	0.42308	0.44050
1.14900	0.24576	0.50343	0.46154	0.52205
1.30550	0.73728	0.45997	0.57692	0.59754
1.46200	0.49152	0.41073	0.65385	0.66572
1.61850	0.24576	0.35930	0.69231	0.72598
1.77500	0.73728	0.30849	0.80769	0.77822
1.93150	0.49152	0.26033	0.88462	0.82268
2.08800	0.00000	0.21618	0.88462	0.85991
2.24450	0.24576	0.17681	0.92308	0.89060
2.40100	0.00000	0.14255	0.92308	0.91552
2.55750	0.00000	0.11335	0.92308	0.93548
2.71400	0.00000	0.08896	0.92308	0.95125
2.87050	0.00000	0.06894	0.92308	0.96355
3.02700	0.24576	0.05278	0.96154	0.97303
3.18350	0.00000	0.03993	0.96154	0.98024
3.34000	0.00000	0.02986	0.96154	0.98567

Figure 2. Sample output file from VTFIT

USING VTFIT

Installing VTFIT

VTFIT can be run from the distribution diskette. However, it runs much faster from a hard disk. To manually install VTFIT on your hard disk:

1. Create a subdirectory for VTFIT on your hard disk, e.g.

   md \vtfit <enter>

2. Copy all the files from the distribution diskette to the VTFIT subdirectory, e.g.

   copy a:*.* c:\vtfit*.* <enter>

3. If you want to be able to run VTFIT from any subdirectory, modify the path in your CONFIG.SYS file to include the VTFIT subdirectory.

Creating input file

While the data to which a distribution is to be fitted may be entered interactively from the keyboard, this practice is not recommended. Rather, an input data file should be created beforehand, using a text editor or a word processor that can save files in ASCII format. This recommended practice allows you to check the data for accuracy. VTFIT does **not** include interactive data editing procedures. In addition the features of the input file mentioned in Chapter 3,

- Columns should be separated by a comma or by two or more blank spaces.
- The index variable may contain any alphanumeric character.
- If an index variable is specified, it has to be comma-delimited from the next column.
- If the index variable contains blank spaces, it has to be enclosed in quotation marks (").

Starting VTFIT

To start VTFIT type

vtfit <enter>

from any subdirectory that your path allows. If you have not modified your path to include the VTFIT subdirectory, then the program can only be run from the subdirectory in which it was placed.

After you start VTFIT you will be presented with three introductory screens. These screens include a statement of the scope of VTFIT and information from the acknowledgments and disclaimer sections in Chapter 1. After you have read each screen, press any key to move forward through the routine.

Specifying output file

The screen that appears after the introductory screens prompts you for the name of file in which you wish the results of your analysis to be placed (Figure 3).

```
Enter the  name of  a  file where output can
be directed. Results will be appended to the
end of this file  to  prevent any  data loss

Press `RETURN' after entering this value"
```

Figure 3. Prompt screen for output file

The name of the output file should include the complete file path. If the file path is omitted, the output file will be placed in the default subdirectory, that is, the subdirectory from which VTFIT is being run. If you simply press the RETURN key, without entering a name, the output will be directed to a file named VTFIT.OUT, located in the default subdirectory. If this file already exists it will be overwritten. This is the only circumstance in which an output file is overwritten. In all other circumstances, the output is appended to the end of existing files.

Specifying input files

After specifying the output file, you will be asked whether or not you have an existing input data file. If not, you are allowed to enter the data that you wish to fit a distribution to. Once you enter a value it cannot be changed. It is therefore better to create an input file before running VTFIT.

If you indicate that you have an existing data file, you will be asked to provide the file specifications indicated in Figure 4.

```
                        DATA FILE SPECIFICATIONS
==================================================================
 1  FILE NAME  (example :- myfile.date.dat)             :file.dat
 2  FILE PATH  (example :- c:\data\)                    :c:\vtfit
 3  USE LEFTMOST COLUMN AS AN INDEX VARIABLE? (Y/N)     :Y
 4  NUMBER OF VARIABLES IN FILE (not counting index var.):3
 5  VARIABLE TO BE FITTED (Counting from left to right)  :1
```

Figure 4. Prompt screen for input file

You are allowed to alter the file specifications until you indicate that you are satisfied with the current set. If you do not specify a file path the file is assumed to be located in the default subdirectory. If the number of the variable to be fitted is greater than the number of variables you specified in the file, you will be asked to change the file specifications. If the file you specified cannot be located in the specified path the routine will be aborted. If this occurs you should recheck the name and location of the file you specified and restart VTFIT. Be sure that the number of variables you specified in the file is correct, or you will get erroneous results.

VTFIT can handle only up to 1000 observations of a variable. In files with more observations, only the first 1000 will be used in the analysis.

Many distributions are bounded by zero at their lower end. If VTFIT encounters an observation that is zero or is negative, you will be prompted to provide a small positive number to be used as the lower limit for those distributions bounded at zero. All observations less than this value will be discarded when VTFIT fits a distribution which is bounded at zero. In the output, the fact that such a distribution has been fitted to a censored subset of the input data, is indicated by reporting both the lower bound and the number of discarded observations.

Fitting a distribution

After the input data has been read into storage, you are given a list of the distributions supported by VTFIT and asked to make a selection (See Figure 5).

Select distribution to be fitted

Gaussian (Normal)
Log Gaussian (Log Normal)
Three parameter Log Gaussian (Log Normal)
Exponential
Shifted Exponential
Beta
Gamma
Three parameter Gamma (Pearson type III)
Log Pearson type III (Log Gamma)
Inverted Gamma (Pearson type V)
Gumbel (Extreme value type I) for minima
Gumbel (Extreme value type I) for maxima
Frechet (Extreme value type II) for minima
Frechet (Extreme value type II) for maxima
Three parameter Frechet for maxima
Weibull (Extreme value type III) for maxima
Weibull (Extreme value type III) for minima
Three parameter Weibull for minima
QUIT

Figure 5. Menu for selecting a distribution

You can scroll through the list of distributions by using the <up arrow> or <down arrow> keys, or by repeatedly pressing the key for the first letter in the name of the distribution to be fitted. After the required distribution is highlighted, press the <enter> key. If a mouse is installed on your system, you can also make a selection by moving the mouse cursor to the required distribution, and double clicking the right mouse button.

After you have selected a distribution, the routine returns the maximum likelihood parameters for that distribution and the corresponding goodness-of-fit statistics. You then have the option of saving or discarding the results. The routine then returns to the distribution fitting menu. You can either choose another distribution in the list or select the QUIT option to move on.

In rare occasions, if you choose a distribution that is an unlikely candidate for the data to which it is being fitted, the routine will be aborted. If, for example, you try to fit the two parameter log Gaussian distribution to a set of negative values, the routine will be aborted due to an *illegal function call* or a *divide by zero* error. If the program is aborted you can simply restart and avoid the particular combination of distribution and data that caused the problem. In some cases it may be worthwhile to transform the data by the addition of a constant before using VTFIT again. It is recommended that you first select the Gaussian distribution to get a visual image of the skew of the data, and try to fit distributions with similar skew (if discernible). This practice reduces the incidences of the routine being aborted.

After you are finished with one variable, you can fit another variable in the same data file, move on to another data file, or you can quit.

REFERENCES

Arora, K. and V.P. Singh. 1989. Comparative evaluation of the estimators of the log Pearson type (LP) 3 distribution. *Journal of Hydrology* 105, 19-37.

Canfield, R.V., *et al.* (1981) Extreme value theory with applications to hydrology. In *Statistical distributions in scientific work, Vol. 6*, ed. C. Taillie, G.P. Patil and B.A. Baldessari, Dordrecht, Holland, D. Reidel Publishing Company, 337-350.

Castillo, E. 1988. *Extreme value theory in engineering.* San Diego, CA: Academic Press, Inc.

Chandra, K.C. 1981. Chi-square goodness of fit tests based on dependent observations. In *Statistical distributions in scientific work, Vol. 5*, ed. C. Taillie, G.P. Patil and B.A. Baldessari, 35-49. Dordrecht, Holland: D. Reidel Publishing Company.

Haan, C.T. 1977. *Statistical methods in hydrology.* Ames, IA: Iowa State University Press.

Kline, D.E. and D.A. Bender. 1990. Maximum likelihood estimation for shifted Weibull and lognormal distributions. *Transactions of the ASAE* 33(1): 330-335.

Kline, D.E. and D.A. Bender. 1990. Maximum likelihood estimation for shifted Weibull and lognormal distributions. *Transactions of the ASAE* 33(1): 330-335.

Law, A.M. and W.D. Kelton. 1991. *Simulation modeling and analysis*. New York: McGraw-Hill, Inc.

Lowing, M.J. 1987. *Casebook of methods for computing hydrological parameters for water projects*. Paris, France, United Nations Educational, Scientific and Cultural Organization.

Rosenbrock, H.H. and C. Storey. 1966. *Computational techniques for chemical engineers*. New York: Pergamon Press.

Shiparo, S.S. and C.W. Brain. 1981. A review of distributional testing procedures and development of a censored sample distributional test. In *Statistical distributions in scientific work,Vol. 5,* ed. C. Taillie, G.P. Patil and B.A. Baldessari, 1-24. Dordrecht, Holland: D. Reidel Publishing Company.

Woeste, F.E., S.K. Suddarth and W.L. Galligan. 1979. Simulation of correlated lumber properties data- A regression approach. *Wood Science* 12(2): 73-79.

Worley, J.W., J.A. Bollinger, F.E. Woeste and K.S. Kline. 1990. Graphic Distribution Analysis (GDA). *Applied Engineering in Agriculture* 6(3): 367-371.