Table 11.5. Empirical coefficients for Δr_{mean} in the MIR criterion.

Distribution	a	b	c
FT-II ($k = 2.5$)	$-2.470 + 0.015\nu^{3/2}$	$-0.1530 - 0.0052\nu^{5/2}$	0
FT-II ($k = 3.33$)	$-2.462 - 0.009\nu^2$	$-0.1933 - 0.0037\nu^{5/2}$	-0.007
FT-II ($k = 5.0$)	-2.463	$-0.2110 - 0.0131\nu^{5/2}$	-0.019
FT-II ($k = 10.0$)	$-2.437 + 0.028\nu^{5/2}$	$-0.2280 - 0.0300\nu^{5/2}$	-0.033
FT-I	$-2.364 + 0.54\nu^{5/2}$	$-0.2665 - 0.0457\nu^{5/2}$	-0.044
Weibull ($k = 0.75$)	$-2.435 - 0.168\nu^{1/2}$	$-0.2083 + 0.1074\nu^{1/2}$	-0.047
Weibull ($k = 1.0$)	-2.355	-0.2612	-0.043
Weibull ($k = 1.4$)	$-2.277 + 0.056\nu^{1/2}$	$-0.3169 - 0.0499\nu$	-0.044
Weibull ($k = 2.0$)	$-2.160 + 0.113\nu$	$-0.3788 - 0.0979\nu$	-0.041
Log-normal	$-2.153 + 0.059\nu^2$	$-0.2627 - 0.1716\nu^{1/4}$	-0.045

For the case of the Kodiak data set with $N = 78$ and $\nu = 1$, Δr_{mean} and $\Delta r/\Delta r_{\text{mean}}$ are calculated for the four distributions of Table 11.4 as follows:

$$
\begin{aligned}
&\text{FT-II } (k = 10) &&: \Delta r_{\text{mean}} = 0.01562, && \Delta r/\Delta r_{\text{mean}} = 0.808\,,\\
&\text{FT-I} &&: \Delta r_{\text{mean}} = 0.01105, && \Delta r/\Delta r_{\text{mean}} = 0.732\,,\\
&\text{Weibull } (k = 1.4) &&: \Delta r_{\text{mean}} = 0.00952, && \Delta r/\Delta r_{\text{mean}} = 0.390\,,\\
&\text{Weibull } (k = 2.0) &&: \Delta r_{\text{mean}} = 0.00743, && \Delta r/\Delta r_{\text{mean}} = 1.472\,.
\end{aligned}
$$

Thus, the Weibull distribution with $k = 1.4$ also satisfies the MIR criterion for best fitting.

The criteria of the largest correlation coefficient and the MIR both give the same judgment of best fitting for the Kodiak data set. However, use of the MIR criterion for the situation where the true parent distribution is unknown requires some caution. In a joint study by several hydraulic institutions on extreme wave statistics,[5] 500 samples of numerically simulated data from the population of the Weibull distribution with $k = 1.4$ were analyzed by various methods. With the least squares method, use of a simple criterion of the largest correlation coefficient produced slightly better results in predicting return values than that of the MIR criterion; further examination by numerical simulation is needed.

(B) *Outlier detection by the DOL criterion*

A sample of extreme data sometimes contains a data which exhibits the value much larger than the rest of data. When the sample is tried to fit to a candidate distribution, the particular data would be plotted at a position far above the line of a fitted distribution curve. Such a data is called an *outlier*. In other cases, the largest data $x_{(1)}$ might be only slightly greater than the second largest data $x_{(2)}$. The data $x_{(1)}$ would then be plotted far below the fitted distribution curve. It is also an outlier.

Detection of an outlier can be made with the DOL (Deviation of OutLier) criterion proposed by Goda and Kobune[6] and/or a statistical test by Barnett and Lewis (Ref. 12, pp. 144–150). The DOL criterion uses the following dimensionless deviation ξ:

$$\xi = \frac{x_{(1)} - \overline{x}}{s}, \tag{11.19}$$

where $\overline{x}$ is the mean of a sample and s is the standard deviation of a sample defined as $s^2 = \sum_{i=1}^{N}(x_i - \overline{x})^2/N$. For the Kodiak data set, $x_{(1)} = 11.7$ m, $\overline{x} = 7.501$ m, $s = 1.206$ m, and thus $\xi =$ is 3.48.

In the statistical test of the normality of a sample, Thompson's test is used by comparing its mean value with the overall mean of a large number of samples. This test can be modified to yield the theoretical value of ξ having the nonexceedance probability P as below.

$$\xi_P = \left[\frac{(N-1)\,F(1, N-2:\alpha)}{N-2+F(1,N-2:\alpha)}\right]^{1/2}, \tag{11.20}$$

where N is the sample size and $F(1, N-2:\alpha)$ denotes the F distribution with the $(1, N-2)$ degrees of freedom at the exceedance probability α. For the largest data $x_{(1)}$, the probability α is given as $2(1 - P^{1/N})$.

The cumulative distribution of ξ has been calculated by Eq. (11.20) and compared with the simulation data sampled from a population of the normal distribution.[6] As shown in Fig. 11.5, the ξ value by simulation agrees with the theory except for the range of low probability. The agreement supports the validity of ξ as a statistical variate. Then the ξ value can be used to judge whether the largest data $x_{(1)}$ of a sample is an outlier or not.

For example, if the ξ value of a sample exceeds the population value $\xi_{95\%}$ corresponding to the exceedance probability of 0.95, the largest data $x_{(1)}$ is judged as an outlier at the level of significance of 0.05. If the ξ value of a sample

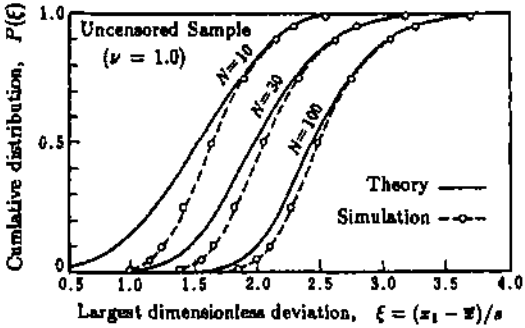

Fig. 11.5. Cumulative distribution of dimensionless deviation of the largest data ξ for samples from the normal distribution.[6]

is below the population value $\xi_{5\%}$ corresponding to the exceedance probability of 0.05, $x_{(1)}$ is also judged as an outlier at the level of significance of 0.05. The threshold value $\xi_{95\%}$ of the population is called the upper DOL, and the threshold value $\xi_{5\%}$ is called the lower DOL. The upper and lower DOLs of various distributions have been estimated with the simulation data of 10,000 samples for respective conditions, by reading the cumulative distribution curves of simulation data such as those shown in Fig. 11.5 as the function of the sample size N and the censoring parameter ν. The measured ξ values have been approximated by the following empirical formula:

$$\xi_{95\%} \text{ and } \xi_{5\%} = a + b \ln N + c (\ln N)^2 \,. \tag{11.21}$$

The empirical coefficients a, b and c have been formulated as listed in Tables 11.6 and 11.7. Relative error of Eq. (11.21) in predicting $\xi_{95\%}$ and $\xi_{5\%}$ is less than $\pm 2\%$.

The DOL criterion for outlier detection is applicable for any sample, and it is independent of distribution fitting methods being employed. It simply determines whether the largest data of a sample is an outlier of a distribution function fitted to a sample or not. In the case of the Kodiak data set, the largest data of 11.7 m is not an outlier for nine distributions discussed in this chapter. When the largest data of a sample is judged as an outlier, a quality check of the data should be made first. If no error is found in the data acquisition process, then the data should not be removed from a sample but the distribution fitted to the sample should be eliminated from the candidate distributions instead.

Table 11.6. Empirical coefficients for the upper DOL criterion $\zeta_{95\%}$.

Distribution	a	b	c
FT-II ($k = 2.5$)	$4.653 - 1.076\nu^{1/2}$	$-2.047 + 0.307\nu^{1/2}$	0.635
FT-II ($k = 3.33$)	$3.217 - 1.216\nu^{1/4}$	$-0.903 + 0.294\nu^{1/4}$	0.427
FT-II ($k = 5.0$)	$0.599 - 0.038\nu^2$	$0.518 - 0.045\nu^2$	0.210
FT-II ($k = 10.0$)	$-0.371 + 0.171\nu^2$	$1.283 - 0.133\nu^2$	0.045
FT-I	$-0.579 + 0.468\nu$	$1.496 - 0.227\nu^2$	-0.038
Weibull ($k = 0.75$)	$-0.256 - 0.632\nu^2$	$1.269 + 0.254\nu^2$	0.037
Weibull ($k = 1.0$)	-0.682	1.600	-0.045
Weibull ($k = 1.4$)	$-0.548 + 0.452\nu^{1/2}$	$1.521 - 0.184\nu$	-0.065
Weibull ($k = 2.0$)	$-0.322 + 0.641\nu^{1/2}$	$1.414 - 0.326\nu$	-0.069
Log-normal	$0.178 + 0.740\nu$	$1.148 - 0.480\nu^{3/2}$	-0.035

Table 11.7. Empirical coefficients for the lower DOL criterion $\xi_{5\%}$.

Distribution	a	b	c
FT-II ($k = 2.5$)	$1.481 - 0.126\nu^{1/4}$	$-0.331 - 0.031\nu^2$	0.192
FT-II ($k = 3.33$)	1.025	$-0.077 - 0.050\nu^2$	0.143
FT-II ($k = 5.0$)	$0.700 + 0.060\nu^2$	$0.139 - 0.076\nu^2$	0.100
FT-II ($k = 10.0$)	$0.424 + 0.088\nu^2$	$0.329 - 0.094\nu^2$	0.061
FT-I	$0.257 + 0.133\nu^2$	$0.452 - 0.118\nu^2$	0.032
Weibull ($k = 0.75$)	$0.534 - 0.162\nu$	$0.277 + 0.095\nu$	0.065
Weibull ($k = 1.0$)	0.308	0.423	0.037
Weibull ($k = 1.4$)	$0.192 + 0.126\nu^{3/2}$	$0.501 - 0.081\nu^{3/2}$	0.018
Weibull ($k = 2.0$)	$0.050 + 0.182\nu^{3/2}$	$0.592 - 0.139\nu^{3/2}$	0
Log-normal	$0.042 + 0.270\nu$	$0.581 - 0.217\nu^{3/2}$	0

(C) *Rejection of candidate distribution by the REC criterion*

Presence of an outlier suggests that a particular distribution is better eliminated from the candidates of parent distributions. When the distribution fitting is made with the least squares method, the value of the correlation coefficient r between the ordered variate $x_{(m)}$ and the reduced variate $y_{(m)}$ can provide another test for rejection of candidate distributions. For this purpose, the residue of correlation coefficient from 1; i.e., $\Delta r = 1 - r$, is employed.

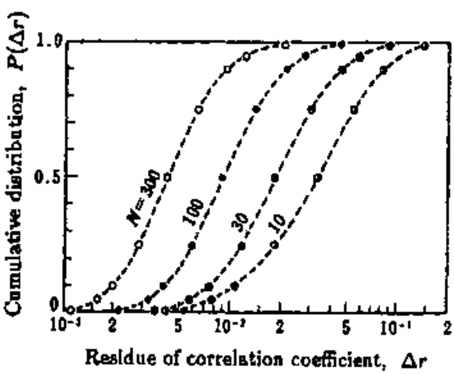

Fig. 11.6. Cumulative distribution of the residue of correlation coefficient Δr for uncensored samples from the Weibull distributions with $k = 1.0$.[6]

Figure 11.6 shows the cumulative distribution curves of Δr for the Weibull distribution with $k = 1$ for the sample size of 10 to 300, which were prepared from simulation data of 10,000 samples for respective conditions.[6] By assuming these cumulative distributions of simulated data being almost the same as those of the population, a criterion for the rejection of candidate function which is called the REC (REsidue of Correlation coefficient) has been prepared. The exceedance probability of 0.95 was set for establishing the threshold value of Δr at the level of significance of 0.05. The threshold value $\Delta r_{95\%}$ has been obtained from the simulation data and formulated in the following empirical expression of Eq. (11.22) with the coefficients listed in Table 11.8.[6] Relative error in predicting the threshold value $\Delta r_{95\%}$ is mostly less than $\pm 3\%$.

$$\Delta r_{95\%} = \exp[a + b \ln N + c (\ln N)^2] . \tag{11.22}$$

In the case of the Kodiak data, all the nine candidate distribution functions yield the residual correlation coefficient Δr below the threshold value $\Delta r_{95\%}$, and thus they are not rejected from the candidates of the parent distribution.

11.3 Estimation of Return Value and Its Confidence Interval

11.3.1 *Statistical Variability of Samples of Extreme Distributions*

In the ocean environment, we sometimes encounter the event of extremely large storm waves having the return period far exceeding the design condition. It could have been generated by a truly abnormal meteorological condition, but

Table 11.8. Empirical Coefficients for $\Delta r_{95\%}$ in the REC criterion.

Distribution	a	b	c
FT-II ($k = 2.5$)	$-1.122 - 0.037\nu$	$-0.3298 + 0.0105\nu^{1/4}$	0.016
FT-II ($k = 3.33$)	$-1.306 - 0.105\nu^{3/2}$	$-0.3001 + 0.0404\nu^{1/2}$	0
FT-II ($k = 5.0$)	$-1.463 - 0.107\nu^{3/2}$	$-0.2716 + 0.0517\nu^{1/4}$	-0.018
FT-II ($k = 10.0$)	$-1.490 - 0.073\nu$	$-0.2299 - 0.0099\nu^{5/2}$	-0.034
FT-I	-1.444	$-0.2733 - 0.0414\nu^{5/2}$	-0.045
Weibull ($k = 0.75$)	$-1.473 - 0.049\nu^2$	$-0.2181 + 0.0505\nu$	-0.041
Weibull ($k = 1.0$)	-1.433	-0.2679	-0.044
Weibull ($k = 1.4$)	-1.312	$-0.3356 - 0.0449\nu$	-0.045
Weibull ($k = 2.0$)	$-1.188 + 0.073\nu^{1/2}$	$-0.4401 - 0.0846\nu^{3/2}$	-0.039
Log-normal	$-1.362 + 0.360\nu^{1/2}$	$-0.3439 - 0.2185\nu^{1/2}$	-0.035

it often results from the situation such that the extreme distribution of storm waves estimated from the previous data set was inappropriate because the sample size was not large enough. This is the problem of statistical variability of samples of extreme statistics.

Figure 11.7 provides an example of variation of samples drawn from an extreme distribution.[5] One hundred samples with the size 100 are numerically simulated from the population of the Weibull distribution with $k = 1.4$, which has the 1-year wave height of 8 m and the 100-year wave height of 13 m at the mean rate $\lambda = 5$; the sample size 100 is equivalent to the period of $K = 20$ years

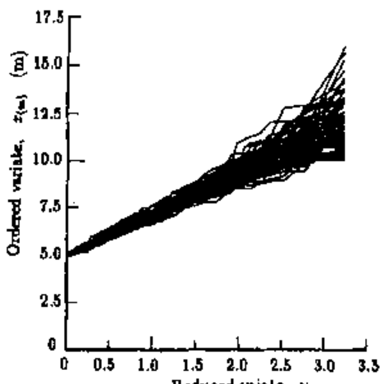

Fig. 11.7. Plot of 100 samples of simulated storm wave height data.[5]

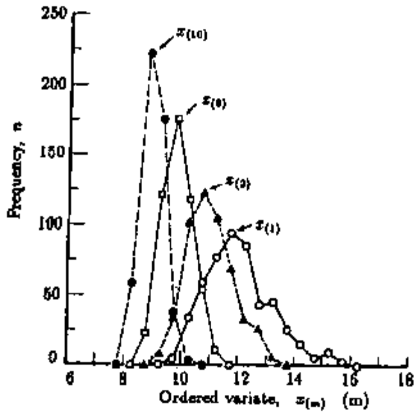

Fig. 11.8. Histograms of the first, second, fifth and tenth largest data of 500 samples.[5]

in duration. As each line in Fig. 11.7 represents the plot of $x_{(m)}$ versus $y_{(m)}$ of one sample, tbe spread of lines indicate the extent of sample variability. Figure 11.8 shows the histograms of the first, second, fifth and tenth largest wave beights of 500 samples drawn from the same distribution. The largest wave height in 20 years varies from 9 to 16 m, even though the 100-year wave height of the population is 13 m. The data set was another subject of joint examination by the working group of the Section of Maritime Hydraulics of IAHR.

The extent of sample variability can be examined through the analysis of the standard deviation of a sample. Although the unbiased variance of a sample defined as $\sigma_x^2 = \sum_{i=1}^{N}(x_i - \bar{x})^2/(N-1)$ has the expected value equal to the population variance, the expected value of σ_x itself is smaller than the population standard deviation as listed in Table 11.9. The difference increases as the sample size decreases. Thus, the terminology of *unbiased standard deviation* should be used with some caution.

Analysis of the statistical characteristics of sample (unbiased) standard deviation has been made for various extreme distribution functions by the author (Ref. 4 and an unpublished source) through a Monte Carlo simulation technique. Table 11.9 lists the mean and the coefficient of variation of the unbiased standard deviation of a sample of extreme distributions. The data is based on the analysis of 20,000 simulated samples for each case, and the results have been slightly modified so that a smooth variation with respect to the sample size N would be realized.

Table 11.9. Mean and coefficient of variation of the sample standard deviation.

Size	FT-II				FT-I
N	$k = 2.5$	$k = 3.33$	$k = 5.0$	$k = 10.0$	
10	0.632 (1.25)	0.798 (0.79)	0.882 (0.55)	0.928 (0.41)	0.951 (0.32)
14	0.665 (1.12)	0.829 (0.71)	0.906 (0.49)	0.946 (0.36)	0.965 (0.27)
20	0.690 (0.99)	0.851 (0.63)	0.923 (0.42)	0.958 (0.31)	0.975 (0.23)
30	0.714 (0.87)	0.873 (0.54)	0.940 (0.36)	0.969 (0.25)	0.984 (0.19)
40	0.732 (0.80)	0.886 (0.49)	0.949 (0.32)	0.975 (0.22)	0.988 (0.16)
60	0.763 (0.74)	0.905 (0.44)	0.963 (0.28)	0.983 (0.18)	0.992 (0.13)
100	0.793 (0.66)	0.929 (0.38)	0.974 (0.23)	0.989 (0.15)	0.995 (0.10)
140	0.810 (0.61)	0.941 (0.35)	0.981 (0.20)	0.992 (0.13)	0.996 (0.09)
200	0.827 (0.56)	0.951 (0.32)	0.987 (0.18)	0.994 (0.11)	0.998 (0.07)

Size	Weibull				Normal
N	$k = 0.75$	$k = 1.00$	$k = 1.4$	$k = 2.0$	
10	0.870 (0.56)	0.925 (0.42)	0.954 (0.31)	0.971 (0.25)	0.975 (0.24)
14	0.898 (0.49)	0.941 (0.36)	0.967 (0.26)	0.980 (0.20)	0.982 (0.20)
20	0.923 (0.42)	0.958 (0.30)	0.977 (0.22)	0.986 (0.17)	0.987 (0.17)
30	0.944 (0.35)	0.971 (0.25)	0.985 (0.18)	0.991 (0.14)	0.992 (0.14)
40	0.957 (0.31)	0.978 (0.22)	0.989 (0.15)	0.993 (0.12)	0.994 (0.12)
60	0.970 (0.25)	0.984 (0.18)	0.992 (0.13)	0.995 (0.10)	0.996 (0.09)
100	0.980 (0.20)	0.990 (0.14)	0.995 (0.10)	0.997 (0.08)	0.998 (0.07)
140	0.985 (0.18)	0.993 (0.12)	0.996 (0.08)	0.998 (0.06)	0.998 (0.06)
200	0.989 (0.15)	0.995 (0.10)	0.997 (0.07)	0.999 (0.06)	0.999 (0.05)

Note: The figures outside the parentheses denote the ratio of the mean of the sample standard deviation to the population value, and the figures inside the parentheses are the coefficient of variation.

Decrease of the sample standard deviation from the population value causes a consistent underestimation of the scale parameter A when the method of moments is employed. Gumbel (Ref. 3, Sec. 6.2.3) lists a table of the sample standard deviation for various sample sizes for the FT-I distribution, but the table gives the values smaller than those listed in Table 11.9. As the data source of Gumbel's table is unknown, it is safe to refer to Table 11.9 when necessary.

11.3.2 *Confidence Interval of Parameter Estimates*

Scatter of samples around the population such as those shown in Fig. 11.7 suggests a certain scatter of the parameter values fitted to each sample. Thus, the estimates of the parameters of extreme distributions constitute statistical variates having their own distributions. It is important to examine the confidence intervals of the parameter estimates. The magnitude of confidence intervals depends on the method of data fitting to a theoretical distribution. When the method of moments is employed, the variability of scale parameter is the same as that of sample standard deviation. For the case of maximum likelihood method, a few theoretical analyses on the parameter estimates are available. Lawless[13] has given a solution for the estimates of parameters of the FT-I distribution, and Challenor[14] has prepared tables of the confidence intervals of the parameter estimates.

For the case of the least squares method, no theory is available on the parameter estimates. Thus, the author (Ref. 4 and an unpublished source) has analyzed the data of the Monte Carlo simulation studies to estimate the confidence intervals of the scale and location parameters of various distributions. The results are listed in Table 11.10, which is based on the data of 20,000 samples for each case. It is found that the confidence intervals of the parameter estimates by the least squares method for the FT-I distribution are about 20% greater than those by the maximum likelihood method, according to the comparison with Challenor's tables.

Example 11.1

Estimates of $\hat{A} = 1.20$ and $\hat{B} = 4.77$ have been obtained from a sample of the FT-I distribution with the size $N = 20$. What are the 95% confidence intervals of these parameters?

Solution

By reading the row of $N = 20$ of the FT-I distribution of Table 11.10, we have the following dimensionless values at 2.5% and 97.5% levels:

$$[A/\hat{A}]_{2.5\%} = 0.67\,, \quad [A/\hat{A}]_{97.5\%} = 1.61\,,$$
$$[(\hat{B} - B)/\hat{A}]_{2.5\%} = -0.49\,, \quad [(\hat{B} - B)/\hat{A}]_{97.5\%} = 0.53\,.$$

Thus, the confidence intervals of the population parameters are calculated as follows:

$$A = (0.67 \sim 1.61) \times 1.20 = 0.80 \sim 1.93\,,$$
$$B = 4.77 + (-0.49 \sim 0.53) \times 1.20 = 4.18 \sim 5.41\,.$$

Table 11.10. Confidence intervals of parameter estimates by the least squares method.

Distribution	N	Scale parameter $A/\hat{A}$					Location parameter $(\hat{B} - B)/\hat{A}$				
		2.5%	25%	75%	97.5%	σ	2.5%	25%	75%	97.5%	σ
FT-II ($k = 2.5$)	10	0.30	0.89	2.04	4.05	0.98	−1.27	−0.31	0.39	1.42	0.66
	14	0.31	0.89	1.90	3.45	0.82	−0.92	−0.25	0.37	1.21	0.53
	20	0.35	0.89	1.78	3.05	0.70	−0.66	−0.20	0.35	1.05	0.44
	30	0.37	0.89	1.65	2.65	0.59	−0.51	−0.15	0.33	0.92	0.37
	40	0.39	0.89	1.59	2.48	0.53	−0.47	−0.13	0.32	0.85	0.34
	60	0.42	0.89	1.51	2.25	0.47	−0.44	−0.10	0.30	0.75	0.30
	100	0.45	0.90	1.43	2.01	0.40	−0.43	−0.08	0.28	0.65	0.28
	140	0.48	0.90	1.39	1.90	0.36	−0.43	−0.08	0.27	0.60	0.26
	200	0.49	0.90	1.35	1.80	0.33	−0.43	−0.07	0.27	0.56	0.25
FT-II ($k = 3.33$)	10	0.36	0.85	1.73	3.19	0.74	−1.09	−0.28	0.30	1.11	0.54
	14	0.38	0.86	1.62	2.73	0.61	−0.82	−0.23	0.27	0.91	0.43
	20	0.42	0.86	1.53	2.43	0.51	−0.62	−0.19	0.24	0.76	0.35
	30	0.46	0.87	1.43	2.14	0.43	−0.47	−0.15	0.21	0.65	0.28
	40	0.49	0.88	1.38	2.00	0.38	−0.40	−0.13	0.20	0.57	0.25
	60	0.52	0.88	1.32	1.83	0.33	−0.34	−0.10	0.17	0.48	0.21
	100	0.56	0.89	1.26	1.65	0.28	−0.30	−0.09	0.15	0.39	0.18
	140	0.59	0.90	1.23	1.56	0.24	−0.28	−0.08	0.14	0.34	0.16
	200	0.62	0.91	1.20	1.49	0.22	−0.26	−0.07	0.12	0.30	0.14
FT-II ($k = 5.0$)	10	0.43	0.84	1.53	2.63	0.57	−0.96	−0.26	0.26	0.95	0.47
	14	0.46	0.85	1.44	2.29	0.47	−0.74	−0.21	0.22	0.76	0.37
	20	0.50	0.86	1.36	2.03	0.39	−0.57	−0.18	0.19	0.62	0.30
	30	0.55	0.87	1.29	1.81	0.32	−0.44	−0.14	0.16	0.51	0.24
	40	0.59	0.88	1.25	1.70	0.28	−0.37	−0.12	0.15	0.44	0.21
	60	0.62	0.89	1.21	1.57	0.24	−0.30	−0.10	0.12	0.36	0.17
	100	0.67	0.91	1.16	1.44	0.19	−0.24	−0.08	0.09	0.28	0.13
	140	0.71	0.91	1.14	1.37	0.17	−0.21	−0.07	0.08	0.24	0.11
	200	0.74	0.92	1.12	1.32	0.15	−0.18	−0.06	0.07	0.20	0.10
FT-II ($k = 10.0$)	10	0.51	0.84	1.39	2.26	0.46	−0.85	−0.24	0.25	0.86	0.42
	14	0.54	0.86	1.32	1.98	0.37	−0.68	−0.19	0.21	0.69	0.34
	20	0.59	0.87	1.26	1.77	0.30	−0.53	−0.16	0.18	0.55	0.27
	30	0.64	0.88	1.20	1.60	0.24	−0.42	−0.13	0.15	0.44	0.22
	40	0.67	0.90	1.18	1.51	0.21	−0.36	−0.11	0.13	0.38	0.19
	60	0.71	0.91	1.14	1.41	0.18	−0.28	−0.09	0.10	0.31	0.15
	100	0.76	0.92	1.11	1.31	0.14	−0.22	−0.07	0.08	0.23	0.12
	140	0.79	0.93	1.09	1.26	0.12	−0.19	−0.06	0.07	0.20	0.10
	200	0.82	0.94	1.08	1.22	0.10	−0.16	−0.05	0.06	0.16	0.08

Note: σ stands for the standard deviation of respective parameter estimates in dimensionless forms.

Table 11.10. (*Continued*)

Distribution	N	Scale parameter $A/\hat{A}$					Location parameter $(\hat{B}-B)/\hat{A}$				
		2.5%	25%	75%	97.5%	σ	2.5%	25%	75%	97.5%	σ
FT-I	10	0.58	0.86	1.30	2.03	0.37	−0.77	−0.21	0.25	0.83	0.40
	14	0.62	0.87	1.24	1.78	0.30	−0.61	−0.18	0.21	0.66	0.32
	20	0.67	0.88	1.19	1.61	0.24	−0.49	−0.15	0.18	0.53	0.26
	30	0.71	0.90	1.14	1.47	0.19	−0.39	−0.12	0.14	0.42	0.20
	40	0.74	0.91	1.13	1.39	0.17	−0.34	−0.11	0.12	0.36	0.18
	60	0.78	0.92	1.10	1.30	0.13	−0.27	−0.09	0.10	0.29	0.14
	100	0.82	0.94	1.08	1.23	0.10	−0.21	−0.07	0.07	0.22	0.11
	140	0.85	0.95	1.06	1.19	0.09	−0.18	−0.06	0.06	0.19	0.09
	200	0.87	0.95	1.05	1.16	0.07	−0.15	−0.05	0.05	0.18	0.08
Weibull ($k=0.75$)	10	0.42	0.82	1.70	3.58	0.86	−0.41	−0.12	0.36	1.14	0.40
	14	0.47	0.83	1.55	2.88	0.65	−0.39	−0.11	0.31	0.93	0.34
	20	0.50	0.84	1.43	2.45	0.50	−0.36	−0.11	0.26	0.75	0.29
	30	0.56	0.85	1.34	2.09	0.40	−0.33	−0.10	0.22	0.61	0.24
	40	0.59	0.87	1.28	1.89	0.33	−0.31	−0.09	0.19	0.53	0.21
	60	0.64	0.88	1.22	1.67	0.27	−0.28	−0.08	0.16	0.42	0.18
	100	0.70	0.90	1.17	1.50	0.20	−0.24	−0.07	0.12	0.33	0.15
	140	0.73	0.91	1.15	1.41	0.18	−0.22	−0.06	0.11	0.28	0.13
	200	0.77	0.92	1.12	1.34	0.15	−0.19	−0.06	0.09	0.24	0.11
Weibull ($k=1.0$)	10	0.51	0.83	1.44	2.58	0.55	−0.34	−0.12	0.24	0.83	0.30
	14	0.55	0.84	1.36	2.22	0.43	−0.31	−0.11	0.21	0.66	0.25
	20	0.60	0.86	1.28	1.92	0.34	−0.28	−0.09	0.17	0.52	0.21
	30	0.65	0.88	1.22	1.70	0.26	−0.25	−0.08	0.13	0.41	0.16
	40	0.69	0.89	1.18	1.58	0.23	−0.22	−0.08	0.12	0.35	0.15
	60	0.72	0.90	1.15	1.44	0.18	−0.20	−0.07	0.10	0.28	0.12
	100	0.78	0.92	1.11	1.33	0.14	−0.16	−0.05	0.07	0.21	0.09
	140	0.80	0.93	1.09	1.27	0.12	−0.14	−0.05	0.06	0.18	0.08
	200	0.83	0.94	1.08	1.22	0.10	−0.12	−0.04	0.05	0.15	0.07
Weibull ($k=1.4$)	10	0.60	0.86	1.30	2.05	0.38	−0.30	−0.12	0.17	0.66	0.24
	14	0.64	0.87	1.24	1.79	0.30	−0.27	−0.10	0.15	0.52	0.20
	20	0.69	0.88	1.19	1.61	0.24	−0.24	−0.09	0.11	0.38	0.15
	30	0.73	0.90	1.14	1.47	0.19	−0.20	−0.07	0.10	0.31	0.13
	40	0.75	0.91	1.12	1.39	0.16	−0.18	−0.07	0.08	0.26	0.11
	60	0.79	0.92	1.10	1.29	0.13	−0.15	−0.06	0.07	0.21	0.09
	100	0.83	0.94	1.07	1.22	0.10	−0.12	−0.05	0.05	0.16	0.07
	140	0.86	0.95	1.06	1.18	0.08	−0.11	−0.04	0.04	0.13	0.06
	200	0.88	0.96	1.05	1.15	0.07	−0.09	−0.03	0.04	0.11	0.05

Note: σ stands for the standard deviation of respective parameter estimates in dimensionless forms.

Table 11.10. (*Continued*)

Distribution	N	Scale parameter $A/\hat{A}$					Location parameter $(\hat{B}-B)/\hat{A})$				
		2.5%	25%	75%	97.5%	σ	2.5%	25%	75%	97.5%	σ
Weibull	10	0.66	0.87	1.22	1.80	0.29	−0.30	−0.12	0.17	0.66	0.24
($k = 2.0$)	14	0.70	0.88	1.17	1.60	0.23	−0.26	−0.10	0.13	0.49	0.19
	20	0.74	0.90	1.14	1.46	0.19	−0.22	−0.09	0.11	0.38	0.15
	30	0.78	0.92	1.11	1.36	0.15	−0.19	−0.07	0.08	0.29	0.12
	40	0.80	0.92	1.09	1.29	0.13	−0.17	−0.06	0.07	0.24	0.10
	60	0.83	0.94	1.07	1.23	0.10	−0.14	−0.05	0.06	0.19	0.08
	100	0.87	0.95	1.05	1.17	0.08	−0.11	−0.04	0.04	0.14	0.06
	140	0.89	0.96	1.05	1.14	0.06	−0.10	−0.04	0.04	0.12	0.05
	200	0.90	0.97	1.04	1.11	0.05	−0.08	−0.03	0.03	0.09	0.04
Normal	10	0.66	0.87	1.21	1.78	0.29	−0.68	−0.22	0.22	0.71	0.35
	14	0.70	0.89	1.16	1.58	0.22	−0.57	−0.18	0.18	0.57	0.29
	20	0.75	0.90	1.13	1.44	0.18	−0.46	−0.15	0.15	0.47	0.23
	30	0.79	0.92	1.10	1.33	0.14	−0.37	−0.12	0.12	0.37	0.19
	40	0.81	0.93	1.08	1.28	0.12	−0.32	−0.11	0.11	0.32	0.16
	60	0.84	0.94	1.07	1.21	0.09	−0.26	−0.09	0.09	0.26	0.13
	100	0.87	0.95	1.05	1.16	0.07	−0.20	−0.07	0.07	0.20	0.10
	140	0.89	0.96	1.04	1.13	0.06	−0.17	−0.06	0.06	0.17	0.09
	200	0.91	0.97	1.03	1.10	0.05	−0.14	−0.05	0.05	0.14	0.07

Note: σ stands for the standard deviation of respective parameter estimates in dimensionless forms.

11.3.3 *Return Value and Its Confidence Interval*

(A) *Estimation of return value*

Once the most probable parent distribution is obtained as the distribution best fitting to the sample under analysis, the return value for a given return period is estimated with the following equation:

$$\hat{x}_R = \hat{B} + \hat{A}\, y_R\,, \tag{11.23}$$

where the reduced variate y_R is calculated as a function of the return period R and the mean rate λ as follows:

$$\left.\begin{aligned} &\text{FT-I distribution} &&: y_R = -\ln\left\{-\ln\left[1 - \frac{1}{\lambda R}\right]\right\}\,, \\ &\text{FT-II distribution} &&: y_R = k\left\{\left[-\ln\left(1 - \frac{1}{\lambda R}\right)\right]^{-1/k} - 1\right\}\,, \\ &\text{Weibull distribution} &&: y_R = [\ln(\lambda R)]^{1/k}\,. \end{aligned}\right\} \tag{11.24}$$

In the case of the Kodiak data, the Weibull distribution with $k = 1.4$ has been concluded as the best-fitting distribution in Sec. 11.2.3. The scale and location parameters have been estimated as $\hat{A} = 1.8621$ m and $\hat{B} = 5.805$ m. The 100-year wave height is estimated as follows:

$$y_{100} = [\ln(3.9 \times 100)]^{1/1.4} = 3.5815\,,$$

$$x_{100} = 5.805 + 1.8621 \times 3.5815 = 12.47\,\text{m}\,.$$

Figure 11.9 shows the fitting of the Kodiak data to the Weibull distribution with $k = 1.4$. The solid line represents the best-fitting line, while the dashed line indicates the 90% confidence interval to be discussed below. The scale in the upper axis indicates the return period in years.

(B) *Statistical variability of return value*

A sample of extreme wave data exhibits quite a large magnitude of variability as demonstrated in Figs. 11.7 and 11.8. The return values estimated with the best-fitting distribution function vary considerably around the true values of population. Figure 11.10 shows examples of statistical variability of return values estimated from individual samples. The FT-I distribution is employed as the population, and uncensored samples ($\nu = 1$) are simulated by a Monte Carlo technique. Two cases of the sample size $N = 20$ and $N = 100$ are analyzed. By assuming the mean rate $\lambda = 1$, the sample sizes correspond to the annual maxima data of 20 and 100 years, respectively. For each sample, the scale and location parameters are estimated with the least squares method by assuming the parent distribution being the FT-I. Then the return value corresponding to the period $R = 100$ (years) is calculated for each sample. The estimates of return value are shown in Fig. 11.10 in a form of probability

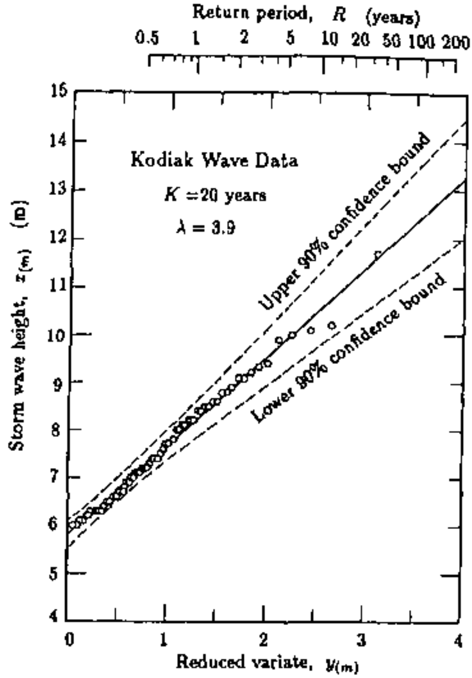

Fig. 11.9. Fitting of the Kodiak storm wave data to the Weibull distribution with $k = 1.4$.

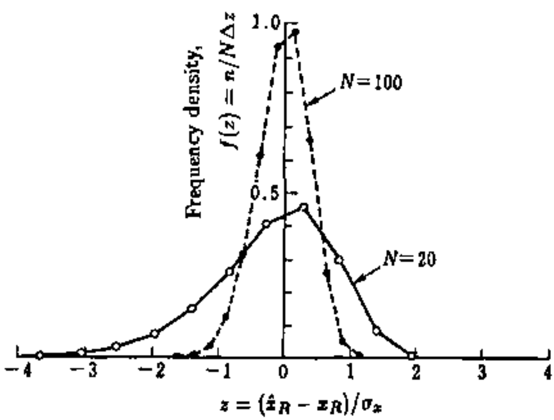

Fig. 11.10. Examples of spread of the return value estimated from uncensored samples of the FT-I distribution: sample sizes $N = 20$ and $N = 100$ for the return period $R = 100$.[17]