

# Calculation method for Hydrologic Statistics

In the case of obtaining the design flood discharge or design probable rainfall, it is necessary to calculate the hydrologic statistics. In this document, formulas and calculation method of the hydrologic statistics are introduced. Please note that the theoretical details are not mentioned in this document.

## Contents

1.	Basic issues	1
1.1	PDF and CDF . . . . .	1
1.2	Return period . . . . .	1
1.3	Average, Variance, Skewness . . . . .	1
1.4	PWM and L-Moments . . . . .	2
1.5	Plotting position formulas . . . . .	2
1.6	Probability Distribution Function for Hydrologic Frequency Analysis . . . . .	3
2.	Probability Distribution Model	4
2.1	Normal distribution (N-distribution) . . . . .	4
2.2	Log-Normal Distribution with 3 parameters (LN3-distribution) . . . . .	5
2.3	Pearson type III distribution (P3-distribution) . . . . .	7
2.4	Log-Pearson type III distribution (LP3-distribution) . . . . .	8
2.5	Gumbel distribution . . . . .	9
2.6	Generalized Extreme Value distribution (GEV-distribution) . . . . .	10
2.7	SQRT exponential-type distribution of maximum (SQRT-ET distribution) . . . . .	11
2.8	Weibull distribution with 3 parameters . . . . .	13
2.9	Exponential distribution . . . . .	15
2.10	Generalized Pareto distribution (GPD) . . . . .	16
3.	Confirmation of fitness	17
3.1	Correlation Coefficient . . . . .	17
3.2	SLSC . . . . .	17
4.	Statistical estimation by Jackknife method	18
5.	Statistical estimation by Bootstrap method	18
6.	Rejection Test	19

## 1. Basic issues

### 1.1 PDF and CDF

The relationship between PDF and CDF is shown below:

$$F(x) = \int_{-\infty}^x f(t)dt \quad (1)$$

where,  $f(x)$  : PDF (probability density function)  
 $F(x)$  : CDF (cumulative distribution function)

According to the definition of CDF, we can understand that  $F(x)$  : CDF is equal to non-exceedance probability for  $x$ .

### 1.2 Return period

#### (1) Return period

Return period is defined as follow:

$$T = \frac{1}{\mu(1-p)} \quad (2)$$

where,  $T$  : return period or recurrence interval (year)  
 $p$  : non-exceedance probability for hydrological variable  $x_p$   
 $p = F(x_p)$        $x_p = F^{-1}(p)$   
 $x_p$  is often called quantile or  $T$ -year event.  
 $\mu$  : annual average number of occurrence of  $X \geq x_p$ .

Above is available for treatment of Peaks Over Threshold data (POT).

If we deal with Annual Maximum Series data (AMS),  $\mu = 1$  :

$$T = \frac{1}{1-p} \quad p = 1 - \frac{1}{T} \quad (3)$$

#### (2) Reliability and Risk for $N$ years (for AMS)

$$P(X < x_T)_N = \left(1 - \frac{1}{T}\right)^N \quad \text{Probability of non-exceedance for } N \text{ years} \quad (4)$$

$$P(X \geq x_T)_N = 1 - \left(1 - \frac{1}{T}\right)^N \quad \text{Probability of exceedance at least once in } N \text{ years} \quad (5)$$

#### (3) Sample calculation (for AMS)

Return period (years)	$T$	5	10	30	100	200	500	1000	5000
Years considered	$N$	5	5	5	10	20	50	100	500
Non-exceedance probability	$P$	0.328	0.590	0.844	0.904	0.905	0.905	0.905	0.905

### 1.3 Average, Variance, Skewness

$$\bar{x} = \frac{1}{N} \sum_{j=1}^N x_j \quad \text{Sample average} \quad (6)$$

$$S^2 = \frac{1}{N} \sum_{j=1}^N (x_j - \bar{x})^2 \quad \text{Sample variance} \quad (7)$$

$$C_s = \frac{1}{N} \sum_{j=1}^N \left(\frac{x_j - \bar{x}}{S}\right)^3 \quad \text{Sample skewness} \quad (8)$$

$$\hat{\sigma}^2 = \frac{N}{N-1} S^2 \quad \text{Unbiased variance} \quad (9)$$

$$\hat{\gamma} = \frac{\sqrt{N(N-1)}}{N-2} C_s \quad \text{Unbiased skewness} \quad (10)$$

#### 1.4 PWM and L-Moments

Probability Weighted Moments (PWM) and L-Moments are adopted in a field of Hydrologic Frequency Analysis.

Although concept of order of data is not included in conventional use of average, variance and skewness, order statistics are adopted in use of PWM and L-Moments.

PWM for a population is defined below:

$$\beta_r = \int_0^1 xF^r dF \quad (r = 0, 1, 2, \dots) \quad (11)$$

and the relationship between PWM and L-Moments for a population is shown below:

$$\lambda_1 = \beta_0 \quad (12)$$

$$\lambda_2 = 2\beta_1 - \beta_0 \quad (13)$$

$$\lambda_3 = 6\beta_2 - 6\beta_1 + \beta_0 \quad (14)$$

PWM for a sample is shown below, and the relationship between PWM and L-Moments is same as the case for population.

$$b_0 = \frac{1}{N} \sum_{j=1}^N x_{(j)} \quad (15)$$

$$b_1 = \frac{1}{N(N-1)} \sum_{j=1}^N (j-1)x_{(j)} \quad (16)$$

$$b_2 = \frac{1}{N(N-1)(N-2)} \sum_{j=1}^N (j-1)(j-2)x_{(j)} \quad (17)$$

where,  $x_{(j)}$  is the  $j$ th value in ascending order.

#### 1.5 Plotting position formulas

A general formula for computing plotting position is defined as follow:

$$F[x_{(i)}] = \frac{i - \alpha}{N + 1 - 2\alpha} \quad (18)$$

where

$N$	Number of sample
$i$	Rank order in ascending order
$x_{(i)}$	Value of sample with rank order $i$ in ascending order
$F[x_{(i)}]$	Plotting position (like a non-exceedance probability)
$\alpha$	Constant for a particular plotting position formula ( $\alpha = 0 \sim 1$ )

Formula	Weibull	Blom	Cunnane	Gringorten	Hazen
$\alpha$	0	0.375	0.40	0.44	0.5

## 1.6 Probability Distribution Function for Hydrologic Frequency Analysis

Some Probability Distribution Functions for Hydrologic Frequency Analysis are enumerated below:

Classification	Name of function	Number of parameters
Distribution of Block Maxima such as	Generalized Extreme Value distribution	3
	Gumbel distribution	2
	Weibull distribution with 3 parameters	3
Annual Maximum Series	SQRT exponential-type distribution of maximum	2
Distribution of Peaks Over Threshold	Generalized Pareto distribution	3
	Exponential distribution	2
Empirical rule	Normal distribution	2
	Log-Normal Distribution with 3 parameters	3
	Pearson type III distribution	3
	Log-Pearson type III distribution	3

## 2. Probability Distribution Model

### 2.1 Normal distribution (N-distribution)

The parameter estimation method in the case of Hydrologic statistic  $x$  following Normal distribution is shown below.

#### (1) Probability Density Function

$$f(x) = \frac{1}{\sigma_x \sqrt{2\pi}} \cdot \exp \left[ -\frac{1}{2} \left( \frac{x - \mu_x}{\sigma_x} \right)^2 \right] \quad (19)$$

#### (2) Cumulative Distribution Function

$$F(x) = \Phi \left( \frac{x - \mu_x}{\sigma_x} \right) \quad \Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z \exp \left( -\frac{1}{2} t^2 \right) dt \quad (20)$$

#### (3) Quantile $x_p$ for non-exceedance probability $p$

$$z = \frac{x - \mu_x}{\sigma_x} \quad \rightarrow \quad x = \mu_x + \sigma_x z \quad (21)$$

$$x_p = \mu_x + \sigma_x z_p \quad z_p: z \text{ value at } p = \Phi(z) \quad (22)$$

#### (4) Estimation of parameters using L-Moments

$$b_0 = \frac{1}{N} \sum_{j=1}^N x_{(j)} \quad b_1 = \frac{1}{N(N-1)} \sum_{j=1}^N (j-1)x_{(j)} \quad (23)$$

where,  $x_{(j)}$  is the  $j$ th value in ascending order.  $x_1$  means minimum value of  $x_j$

The value of  $\lambda_i$  can be calculated by deeming  $b_i = \beta_i$  in above.

$$\lambda_1 = \beta_0 \quad \lambda_2 = 2\beta_1 - \beta_0 \quad (24)$$

Parameters can be estimated using following relationship.

$$\begin{cases} \mu_x = \lambda_1 \\ \sigma_x = \sqrt{\pi} \lambda_2 \end{cases} \quad (25)$$

## 2.2 Log-Normal Distribution with 3 parameters (LN3-distribution)

The parameter estimation method in the case of Hydrologic statistic  $x$  following Log-Normal distribution with 3 parameters is shown below.

### (1) Probability Density Function

$$f(x) = \frac{1}{(x-a)\sigma_y\sqrt{2\pi}} \cdot \exp\left\{-\frac{1}{2}\left[\frac{\ln(x-a)-\mu_y}{\sigma_y}\right]^2\right\} \quad y = \ln(x-a) \quad (26)$$

### (2) Cumulative Distribution Function

$$F(x) = \Phi\left(\frac{\ln(x-a)-\mu_y}{\sigma_y}\right) \quad \Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z \exp\left(-\frac{1}{2}t^2\right) dt \quad (27)$$

### (3) Quantile $x_p$ for non-exceedance probability $p$

$$z = \frac{\ln(x-a)-\mu_y}{\sigma_y} \quad \rightarrow \quad x = a + \exp(\mu_y + \sigma_y z) \quad (28)$$

$$x_p = a + \exp(\mu_y + \sigma_y z_p) \quad z_p: \text{value of } z \text{ at } p = \Phi(z) \quad (29)$$

### (4) Estimation of parameters (Iwai method or Quantile method)

$$\begin{cases} a = \frac{x_{(1)} \cdot x_{(N)} - x_m^2}{x_{(1)} + x_{(N)} - 2x_m} & x_{(1)} + x_{(N)} - 2x_m > 0 \\ \mu_y = \frac{1}{N} \sum_{j=1}^N \ln(x_j - a) \\ \sigma_y^2 = \frac{1}{N} \sum_{j=1}^N [\ln(x_j - a) - \mu_y]^2 \end{cases} \quad (30)$$

where,  $x_{(1)}$  is the minimum value of sample data,  $x_{(N)}$  is the maximum value of sample data and  $x_m$  is the median of sample data.  $x_{(j)}$  is  $j$ -th value of  $x$  in ascending order.

### (5) Estimation of parameters (Moment method)

$$\bar{x} = \frac{1}{N} \sum_{j=1}^N x_j \quad S_x^2 = \frac{1}{N} \sum_{j=1}^N (x_j - \bar{x})^2 \quad C_{sx} = \frac{1}{N} \sum_{j=1}^N \left(\frac{x_j - \bar{x}}{S_x}\right)^3 \quad (31)$$

$$\mu_x = \bar{x} \quad \sigma_x = [N/(N-1)]^{1/2} S_x \quad \gamma_x = \frac{\sqrt{N(N-1)}}{N-2} C_{sx} \quad (32)$$

#### (Reference)

To get the unbiased skewness  $\gamma_x$ , the following formula for Log-Normal distribution by Bobee and Roubaille is adopted. Note that  $C_{sx}$  to the power of 3 is multiplied by  $B$ .

$$\gamma_x = C_{sx}(A + B \cdot C_{sx}^3) \quad (33)$$

$$\text{where,} \quad A = 1.01 + 7.01/N + 14.66/N^2 \quad B = 1.69/N + 74.66/N^2 \quad (34)$$

The relationship between moments and parameters is shown below:

$$\mu_x = a + \exp(\mu_y) \cdot \exp(\sigma_y^2/2) \quad (35)$$

$$\sigma_x = \exp(\mu_y) \sqrt{\exp(\sigma_y^2)\{\exp(\sigma_y^2) - 1\}} \quad (36)$$

$$\gamma_x = \{\exp(\sigma_y^2) + 2\} \sqrt{\exp(\sigma_y^2) - 1} \quad (37)$$

The parameters which we want to know are  $\sigma_y$ ,  $\mu_y$  and  $a$  in above equations. The third equation in above is expressed as follow,

$$\gamma_x = \{\exp(\sigma_y^2) + 2\} \sqrt{\exp(\sigma_y^2) - 1} \Rightarrow x^3 + 3x^2 - 4 - \gamma_x^2 = 0 \quad \text{where,} \quad x = \exp(\sigma_y^2) \quad (38)$$

The real positive solution for above cubic equation can be obtained by Cardano's method as follow,

$$x = \left(\beta + \sqrt{\beta^2 - 1}\right)^{1/3} + \left(\beta - \sqrt{\beta^2 - 1}\right)^{1/3} - 1 \quad \text{where,} \quad \beta = 1 + \frac{\gamma_x^2}{2} \quad (39)$$

As a result, parameters can be estimated using following equations.

$$\begin{cases} \sigma_y = \sqrt{\ln(x)} \\ \mu_y = \ln\left(\frac{\sigma_x}{\sqrt{x(x-1)}}\right) \\ a = \mu_x - \exp(\mu_y) \cdot \exp(\sigma_y^2/2) \end{cases} : x \text{ is a real positive solution of cubic equation } x^3 + 3x^2 - 4 - \gamma_x^2 = 0$$

**(6) Estimation of parameters (Trial method using single regression analysis)**

Parameters can be obtain using following relationship.

$$z = \frac{\ln(x-a) - \mu_y}{\sigma_y} \Rightarrow z = A \cdot X + B \quad (40)$$

$$X = \ln(x-a) \quad A = \frac{1}{\sigma_y} \quad B = -\frac{\mu_y}{\sigma_y} \quad (41)$$

where,  $z$  is a %-point of standard normal distribution. It is calculated from the non-exceedance probability of  $x$ . Non-exceedance probability can be obtained using Protting position formula.

From above, following relationships are derived.

$$\begin{cases} a : (\text{selected value by trial single regression analysis}) \\ \mu_y = -\frac{B}{A} \\ \sigma_y = \frac{1}{A} \end{cases}$$

$a$  is determind as a value in case that linear line has the maximum correlation coefficient. The procedure to obtain the value of  $a$  is shown below:

- a. Set the initial value of  $a = x_{(1)} - \delta$  ( $\delta$ : small value).
- b. Repeat single regression analysis by reducing the value of  $a$ .
- c. Select the value of  $a$  when linear line has the maximum correlation coefficient.

### 2.3 Pearson type III distribution (P3-distribution)

The parameter estimation method in the case of Hydrologic statistic  $x$  following Pearson type III distribution (Gamma distribution) is shown below.

#### (1) Probability Density Function

$$f(x) = \frac{1}{|a| \cdot \Gamma(b)} \left( \frac{x-c}{a} \right)^{b-1} \cdot \exp\left(-\frac{x-c}{a}\right) \quad a > 0 : c \leq x < \infty \quad (42)$$

$c$  is a location parameter,  $a$  is a scale parameter,  $b$  is a shape parameter.

#### (2) Cumulative Distribution Function

$$F(x) = G\left(\frac{x-c}{a}\right) \quad G(w) = \frac{1}{\Gamma(b)} \int_0^w t^{b-1} \exp(-t) dt \quad (a > 0) \quad (43)$$

#### (3) Quantile $x_p$ for non-exceedance probability $p$

$$w = \frac{x-c}{a} \quad \rightarrow \quad x = c + aw \quad (44)$$

$$x_p = c + aw_p \quad w_p: \text{ Value of } w \text{ at } p = G(w) \quad (45)$$

#### (4) Estimation of parameters (Moment method)

$$\bar{x} = \frac{1}{N} \sum_{j=1}^N x_j \quad S_x^2 = \frac{1}{N} \sum_{j=1}^N (x_j - \bar{x})^2 \quad C_{sx} = \frac{1}{N} \sum_{j=1}^N \left( \frac{x_j - \bar{x}}{S_x} \right)^3 \quad (46)$$

$$\mu_x = \bar{x} \quad \sigma_x = [N/(N-1)]^{1/2} S_x \quad \gamma_x = \frac{\sqrt{N(N-1)}}{N-2} C_{sx} \quad (47)$$

#### (Reference)

To get the unbiased skewness  $\gamma_x$ , the following formula for Pearson type III distribution by Bobee and Robitaille is adopted. Note that  $C_{sx}$  to the power of 2 is multiplied by  $B$ .

$$\gamma_x = C_{sx}(A + B \cdot C_{sx}^2) \quad (48)$$

$$\text{where,} \quad A = 1 + 6.51/N + 20.2/N^2 \quad B = 1.48/N + 6.77/N^2 \quad (49)$$

The relationship between moments and parameters is shown below:

$$\mu_x = c + a \cdot b \quad \sigma_x^2 = a^2 \cdot b \quad \gamma_x = \frac{2a}{|a|\sqrt{b}} \quad (50)$$

As a result, parameters can be estimated using following equations.

$$\begin{cases} b = 4/\gamma_x^2 & (b > 0) \\ a = \sigma_x/\sqrt{b} & (\gamma_x < 0 \rightarrow a = -\sigma_x/\sqrt{b} < 0) \\ c = \mu_x - ab \end{cases} \quad (51)$$

Note that if  $\gamma_x$  is less than 0 ( $\gamma_x < 0$ ),  $a$  shall be less than 0 ( $a < 0$ ) and the value of  $w_p$  is for the probability of  $1-p$ .

If the absolute value of  $\gamma_x$  is small, the same problem as Log-Pearson type III distribution may occur.



## 2.4 Log-Pearson type III distribution (LP3-distribution)

The parameter estimation method in the case of Hydrologic statistic  $x$  following Log-Pearson type III distribution is shown below.

### (1) Probability Density Function

$$f(x) = \frac{1}{|a| \cdot \Gamma(b) \cdot x} \left( \frac{\ln x - c}{a} \right)^{b-1} \cdot \exp\left(-\frac{\ln x - c}{a}\right) \quad a > 0 : \exp(c) < x < \infty \quad (52)$$

$c$  is a location parameter,  $a$  is a scale parameter,  $b$  is a shape parameter.

### (2) Cumulative Distribution Function

$$F(x) = G\left(\frac{\ln x - c}{a}\right) \quad G(w) = \frac{1}{\Gamma(b)} \int_0^w t^{b-1} \exp(-t) dt \quad (a > 0) \quad (53)$$

### (3) Quantile $x_p$ for non-exceedance probability $p$

$$w = \frac{\ln x - c}{a} \quad \rightarrow \quad x = \exp(c + aw) \quad (54)$$

$$x_p = \exp(c + aw_p) \quad w_p: \text{the value of } w \text{ at } p = G(w) \quad (55)$$

### (4) Estimation of parameters (Moment method)

$$y_j = \ln x_j \quad \bar{y} = \frac{1}{N} \sum_{j=1}^N y_j \quad S_y^2 = \frac{1}{N} \sum_{j=1}^N (y_j - \bar{y})^2 \quad C_{sy} = \frac{1}{N} \sum_{j=1}^N \left( \frac{y_j - \bar{y}}{S_y} \right)^3 \quad (56)$$

$$\mu_y = \bar{y} \quad \sigma_y = [N/(N-1)]^{1/2} S_y \quad \gamma_y = \frac{\sqrt{N(N-1)}}{N-2} C_{sy} \quad (57)$$

#### (Reference)

To get the unbiased skewness  $\gamma_x$ , the following formula for Pearson type III distribution by Bobee and Robitaille is adopted. Note that  $C_{sy}$  to the power of 2 is multiplied by  $B$ .

$$\gamma_y = C_{sy}(A + B \cdot C_{sy}^2) \quad (58)$$

$$\text{where,} \quad A = 1 + 6.51/N + 20.2/N^2 \quad B = 1.48/N + 6.77/N^2 \quad (59)$$

The relationship between moments and parameters is shown below:

$$\mu_y = c + a \cdot b \quad \sigma_y^2 = a^2 \cdot b \quad \gamma_y = \frac{2a}{|a|\sqrt{b}} \quad (60)$$

As a result, parameters can be estimated using following equations.

$$\begin{cases} b = 4/\gamma_y^2 & (b > 0) \\ a = \sigma_y/\sqrt{b} & (\gamma_y < 0 \rightarrow a = -\sigma_y/\sqrt{b} < 0) \\ c = \mu_y - ab \end{cases} \quad (61)$$

Note that if  $\gamma_y$  is less than 0 ( $\gamma_y < 0$ ),  $a$  shall be less than 0 and the value of  $w_p$  is for the value of  $1 - p$ .

In the case of small absolute value of  $\gamma_y$ , the value of  $b$  becomes huge so that the converged value of percent point for Gamma distribution can not be obtained. Therefore, if  $b$  becomes huge value, quantile  $x_p$  for non-exceedance probability  $p$  can be calculated using Wilson-Hilferty transformation shown below.

$$x_p = \exp(\mu_y + \sigma_y \cdot K_p) \quad K_p = \frac{2}{\gamma_y} \left( 1 + \frac{\gamma_y z_p}{6} - \frac{\gamma_y^2}{36} \right) - \frac{2}{\gamma_y} \quad (62)$$

where,  $z_p$  is standard normal deviate following  $N(0, 1)$ . If  $b$  is greater than 10,000, it may be better to use Wilson-Hilferty transformation. Even if skewness  $\gamma_y$  is less than 0 ( $\gamma_y < 0$ ), standard normal deviate  $z_p$  shall be for the probability of  $p$ , and it is not necessary to change the value of  $\gamma_y$  to  $|\gamma_y|$  (absolute value). It means just calculated values of  $p$  and  $\gamma_y$  can be used for next calculation.

## 2.5 Gumbel distribution

The parameter estimation method in the case of Hydrologic statistic  $x$  following Gumbel distribution is shown below.

### (1) Probability Density Function

$$f(x) = \frac{1}{a} \exp \left[ -\frac{x-c}{a} - \exp \left( -\frac{x-c}{a} \right) \right] \quad -\infty < x < \infty \quad (63)$$

$c$  is a location parameter,  $a$  is a scale parameter.

### (2) Cumulative Distribution Function

$$F(x) = \exp \left[ -\exp \left( -\frac{x-c}{a} \right) \right] \quad (64)$$

### (3) Quantile $x_p$ for non-exceedance probability $p$

$$p = \exp \left[ -\exp \left( -\frac{x-c}{a} \right) \right] \quad \rightarrow \quad x = c - a \ln[-\ln(p)] \quad (65)$$

$$x_p = c - a \ln[-\ln(p)] \quad (66)$$

### (4) Estimation of parameters (L-moments method)

$$b_0 = \frac{1}{N} \sum_{j=1}^N x_{(j)} \quad b_1 = \frac{1}{N(N-1)} \sum_{j=1}^N (j-1)x_{(j)} \quad (67)$$

where,  $x_{(j)}$  is  $j$ -th value of sample  $x$  in ascending order. It means that  $x_{(1)}$  is the minimum value of sample data and  $x_{(N)}$  is the maximum value of sample data.

The values of  $\lambda_i$  can be calculated using following relationship by deeming  $b_i = \beta_i$ .

$$\lambda_1 = \beta_0 \quad \lambda_2 = 2\beta_1 - \beta_0 \quad (68)$$

Parameters can be estimated using following relationship between L-Moments and parameters.

$$\begin{cases} a = \lambda_2 / \ln 2 \\ c = \lambda_1 - 0.5772a \end{cases} \quad (69)$$

## 2.6 Generalized Extreme Value distribution (GEV-distribution)

The parameter estimation method in the case of Hydrologic statistic  $x$  following Generalized Extreme Value distribution is shown below. Gumbel distribution is the same as Generalized Extreme value distribution with  $k = 0$ .

### (1) Probability Density Function

$$f(x) = \frac{1}{a} \left(1 - k \frac{x-c}{a}\right)^{1/k-1} \cdot \exp \left[ - \left(1 - k \frac{x-c}{a}\right)^{1/k} \right] \quad (k \neq 0) \quad (70)$$

$c$  is a location parameter,  $a$  is a scale parameter,  $k$  is a shape parameter.

### (2) Cumulative Distribution Function

$$F(x) = \exp \left[ - \left(1 - k \frac{x-c}{a}\right)^{1/k} \right] \quad (k \neq 0) \quad (71)$$

### (3) Quantile $x_p$ for non-exceedance probability $p$

$$p = \exp \left[ - \left(1 - k \frac{x-c}{a}\right)^{1/k} \right] \quad \rightarrow \quad x = c + \frac{a}{k} \cdot \{1 - [-\ln(p)]^k\} \quad (72)$$

$$x_p = c + \frac{a}{k} \cdot \{1 - [-\ln(p)]^k\} \quad (73)$$

### (4) Estimation of parameters (L-moments method)

$$b_0 = \frac{1}{N} \sum_{j=1}^N x_{(j)} \quad b_1 = \frac{1}{N(N-1)} \sum_{j=1}^N (j-1)x_{(j)} \quad b_2 = \frac{1}{N(N-1)(N-2)} \sum_{j=1}^N (j-1)(j-2)x_{(j)} \quad (74)$$

where,  $x_{(j)}$  is  $j$ -th value of sample  $x$  in ascending order. It means that  $x_{(1)}$  is the minimum value of sample data and  $x_{(N)}$  is the maximum value of sample data.

The values of  $\lambda_i$  can be calculated using following relationship by deeming  $b_i = \beta_i$ .

$$\lambda_1 = \beta_0 \quad \lambda_2 = 2\beta_1 - \beta_0 \quad \lambda_3 = 6\beta_2 - 6\beta_1 + \beta_0 \quad (75)$$

Parameters can be estimated using following relationship between L-Moments and parameters.

$$\begin{cases} k = 7.8590d + 2.9554d^2 & \text{where, } d = \frac{2\lambda_2}{\lambda_3 + 3\lambda_2} - \frac{\ln(2)}{\ln(3)} \\ a = \frac{k\lambda_2}{(1 - 2^{-k}) \cdot \Gamma(1+k)} \\ c = \lambda_1 - \frac{a}{k} \cdot [1 - \Gamma(1+k)] \end{cases} \quad (76)$$

## 2.7 SQRT exponential-type distribution of maximum (SQRT-ET distribution)

The parameter estimation method in the case of Hydrologic statistic  $x$  following SQRT exponential-type distribution of maximum is shown below.

### (1) Probability Density Function

$$f(x) = \frac{ab}{2} \exp \left[ -\sqrt{bx} - a \left( 1 + \sqrt{bx} \right) \exp \left( -\sqrt{bx} \right) \right] \quad (x \geq 0) \quad (77)$$

### (2) Cumulative Distribution Function

$$F(x) = \exp \left[ -a \left( 1 + \sqrt{bx} \right) \exp \left( -\sqrt{bx} \right) \right] \quad (x \geq 0) \quad (78)$$

### (3) Quantile $x_p$ for non-exceedance probability $p$

$$p = \exp \left[ -a \left( 1 + \sqrt{bx} \right) \exp \left( -\sqrt{bx} \right) \right] = \exp \left[ -a(1 + t_p) \exp(-t_p) \right] \quad (t_p = \sqrt{bx}) \quad (79)$$

$$\rightarrow \quad x = \frac{t_p^2}{b} \quad \ln(1 + t_p) - t_p = \ln \left[ -\frac{1}{a} \ln(p) \right] \quad (80)$$

$$x_p = \frac{t_p^2}{b} \quad \ln(1 + t_p) - t_p = \ln \left[ -\frac{1}{a} \ln(p) \right] \quad (81)$$

■(Reference) Calculation method of  $t_p$

We define the function  $g(t_p)$  as follow:

$$g(t_p) = \ln(1 + t_p) - t_p - \ln \left[ -\frac{1}{a} \ln(p) \right] \quad (82)$$

Next we can get the differential of  $g(t_p)$  as follow:

$$g'(t_p) = \frac{1}{1 + t_p} - 1 \quad (83)$$

So we can know that function  $g(t_p)$  is a monotone decreasing function because the value of  $g'(t_p)$  is negative for valid  $t_p$ .

And we can obtain the value of  $t_p$  at  $g(t_p) = 0$  using Newton-Raphson method shown below.

$$t_{p(n+1)} = t_{p(n)} - \frac{g(t_{p(n)})}{g'(t_{p(n)})} \quad (n): \text{counter for iterative calculation} \quad (84)$$

Regarding an initial value of  $t_p$ , it is convenient to use the value of  $t_p = \sqrt{b \cdot x_{max}}$ , because we want to know the value of  $x_p$  at the heigher range of non-exceedance probability of  $p$ .

### (4) Estimation of parameters (Maximum Likelihood Method)

Parameters  $a$  and  $b$  are determined so that a log-likelihood function  $L$  shown below has the maximum value.

$$\begin{aligned} L(a, b) &= \sum_{j=1}^N \ln f(x_j) \\ &= N \ln a + N \ln b - N \ln 2 - \sum_{j=1}^N \sqrt{bx_j} - a \left[ \sum_{j=1}^N \exp \left( -\sqrt{bx_j} \right) + \sum_{j=1}^N \sqrt{bx_j} \exp \left( -\sqrt{bx_j} \right) \right] \end{aligned} \quad (85)$$

From the condition that partial differential of  $L$  by  $b$  becomes 0,  $a$  can be expressed as a function of  $b$ . We define this temporary value equals to  $a_1$ .

$$\frac{\partial L}{\partial b} = 0 \quad \rightarrow \quad a = \frac{\sum_{j=1}^N \sqrt{bx_j} - 2N}{\sum_{j=1}^N (bx_j) \exp \left( -\sqrt{bx_j} \right)} = a_1 \quad (86)$$

Next, from the condition that partial differential of  $L$  by  $a$  becomes 0,  $a$  can be expressed as a function of  $b$ . We define this temporary value equals to  $a_2$ .

$$\frac{\partial L}{\partial a} = 0 \quad \rightarrow \quad a = \frac{N}{\sum_{j=1}^N \exp(-\sqrt{bx_j}) + \sum_{j=1}^N \sqrt{bx_j} \exp(-\sqrt{bx_j})} = a_2 \quad (87)$$

Since  $L$  has the maximum value at the condition of  $a_1 = a_2$ , the value of  $b$  can be obtained using bisection method, where the equation for bisection method is  $h(b) = a_1(b) - a_2(b) = 0$ . Note that  $a_2 > 0$  is everytime true, but  $b$  shall satisfy the following condition for ensuring  $a_1 > 0$ .

$$a_1 > 0 \quad \rightarrow \quad b > \left( \frac{2N}{\sum_{j=1}^N \sqrt{x_j}} \right)^2 \quad (88)$$

Sample program by C language is shown below.

```

/* Bisection method */
b1=bb; /* Set the value of b1 (condition: a1>0) */
b2=b1+0.5; /* Set the value of b2 as the value of b1+0.5 */
bb=0.5*(b1+b2); /* Set the intermediate value of bb */
f1=FSQR(nd,datax,b1,&a1,&a2); /* Get the value of h(b1) */
f2=FSQR(nd,datax,b2,&a1,&a2); /* Get the value of h(b2) */
ff=FSQR(nd,datax,bb,&a1,&a2); /* Get the value of h(bb) */
do{ /* Loop for iterative calculation */
    if(f1*ff<0.0)b2=bb;
    if(ff*f2<0.0)b1=bb;
    if(ff==0.0)break;
    if(0.0<f1*ff&&0.0<ff*f2){b1=b2;b2=b1+0.5;} /* Technique for searching the solution. */
    bb=0.5*(b1+b2);
    f1=FSQR(nd,datax,b1,&a1,&a2);
    f2=FSQR(nd,datax,b2,&a1,&a2);
    ff=FSQR(nd,datax,bb,&a1,&a2);
}while(0.001<fabs(a1-a2)); /* Criteria for judgement of convergence: |h(b)|<0.001 */

```

## 2.8 Weibull distribution with 3 parameters

The parameter estimation method in the case of Hydrologic statistic  $x$  following Weibull distribution with 3 parameters is shown below.

### (1) Probability Density Function

$$f(x) = \frac{k}{a} \left( \frac{x-c}{a} \right)^{k-1} \exp \left[ - \left( \frac{x-c}{a} \right)^k \right] \quad (k \neq 0) \quad (89)$$

$c$  is a location parameter,  $a$  is a scale parameter,  $k$  is a shape parameter.

### (2) Cumulative Distribution Function

$$F(x) = 1 - \exp \left[ - \left( \frac{x-c}{a} \right)^k \right] \quad (k \neq 0) \quad (90)$$

### (3) Quantile $x_p$ for non-exceedance probability $p$

$$1 - p = \exp \left[ - \left( \frac{x-c}{a} \right)^k \right] \quad \rightarrow \quad x = c + a[-\ln(1-p)]^{1/k} \quad (91)$$

$$x_p = c + a[-\ln(1-p)]^{1/k} \quad (92)$$

### (4) Estimation of parameters (L-moments method by Goda \*)

$$b_0 = \frac{1}{N} \sum_{j=1}^N x_{(j)} \quad b_1 = \frac{1}{N(N-1)} \sum_{j=1}^N (j-1)x_{(j)} \quad b_2 = \frac{1}{N(N-1)(N-2)} \sum_{j=1}^N (j-1)(j-2)x_{(j)} \quad (93)$$

where,  $x_{(j)}$  is  $j$ -th value of sample  $x$  in ascending order. It means that  $x_{(1)}$  is the minimum value of sample data and  $x_{(N)}$  is the maximum value of sample data.

The values of  $\lambda_i$  can be calculated using following relationship by deeming  $b_i = \beta_i$ .

$$\lambda_1 = \beta_0 \quad \lambda_2 = 2\beta_1 - \beta_0 \quad \lambda_3 = 6\beta_2 - 6\beta_1 + \beta_0 \quad (94)$$

Parameters can be estimated using following relationship between L-Moments and parameters.

$$\begin{cases} k = 285.3\tau^6 - 658.6\tau^5 + 622.8\tau^4 - 317.2\tau^3 + 98.52\tau^2 - 21.256\tau + 3.5160 & \text{where } \tau = \lambda_3/\lambda_2 \\ a = \frac{\lambda_2}{(1 - 2^{-1/k}) \cdot \Gamma(1 + 1/k)} \\ c = \lambda_1 - a \cdot \Gamma(1 + 1/k) \end{cases} \quad (95)$$

\*) [https://www.jstage.jst.go.jp/article/kaigan/65/1/65\\_1\\_161/\\_pdf](https://www.jstage.jst.go.jp/article/kaigan/65/1/65_1_161/_pdf)

"Use of L-moments Method for Extreme Statistics of Storm Wave Heights"

(author: Yoshimi GODA, Masanobu KUDAKA and Hiroyasu KAWAI)

For the item of Weibull distribution in the Table-2 on above paper:

Although  $k$  is the function of  $\lambda_3$  in this table, I think  $\tau_3 = \lambda_3/\lambda_2$  is correct. In addition, although denominator of  $A$  includes  $k$ , I think this is miss type.

### (5) Estimation of parameters (Maximum Likelihood Method)

Since 3 parameters are included in this problem, following procedure can be adopted.

- First, to estimate location parameter  $c$  using plotting position formula. As a result,  $c$  becomes known parameter.
- Second, to estimate shape parameter  $k$  and scale parameter  $a$  using known parameter of  $c$ . As a calculation method, Maximum Likelihood Method can be used.

#### (I) Estimation of location parameter $c$

Non-exceedance probability  $F(x)$  of sample  $x$  can be estimated using plotting position formula. Note that sample data shall be sorted order in ascending to use plotting position formula.

Following equation is obtained by transforming Cumulative Distribution Function.

$$1 - F(x) = \exp \left[ - \left( \frac{x - c}{a} \right)^k \right] \quad (96)$$

From above, following equation can be obtained as a logarithm of 2 times.

$$\ln\{-\ln[1 - F(x)]\} = k \ln(x - c) - k \ln a \quad \rightarrow \quad Y = A \cdot X + B \quad (97)$$

Next,  $k$  and  $a$  can be obtained using following equations based on single regression analysis.

$$Y_i = \ln\{-\ln[1 - F(x_i)]\} \quad X_i = \ln(x_i - c) \quad (98)$$

$$A = \frac{N \sum X_i Y_i - \sum X_i \cdot \sum Y_i}{N \sum X_i^2 - (\sum X_i)^2} \quad B = \frac{\sum X_i^2 \cdot \sum Y_i - \sum X_i \cdot \sum X_i Y_i}{N \sum X_i^2 - (\sum X_i)^2} \quad (99)$$

$$r = \frac{N \sum X_i Y_i - \sum X_i \cdot \sum Y_i}{\sqrt{[N \sum X_i^2 - (\sum X_i)^2] \cdot [N \sum Y_i^2 - (\sum Y_i)^2]}} \quad (100)$$

$$k = A \quad a = \exp(-B/A) \quad (101)$$

where,  $N$  is sample number.

Since  $c$  is unknown value in above equations, it is necessary to carry out the trial calculation to get the optimal value of  $c$  which gives maximum value of correlation coefficient  $r$  by simple regression analysis.

In this stage, temporary values of  $k$ ,  $a$  and  $c$  are obtained. However, it is desirable to carry out additional calculation in order to get the more certain values.

In next stage, the condition is that  $c$  is fixed and  $k$  and  $a$  are unknown values.

## (II) Estimation of shape parameter $k$ and scale parameter $a$

By setting  $t = x - c$ , following equation is obtained, where  $c$  is already fixed (known parameter).

$$f(t) = \frac{k}{a} \left( \frac{t}{a} \right)^{k-1} \exp \left[ - \left( \frac{t}{a} \right)^k \right] \quad (102)$$

Using log likelihood function for above equation  $L = \sum_{i=1}^N \ln f(t_i)$ ,  $k$  and  $a$  can be estimated by Newton-Raphson method. As an initial value of  $k$ , obtained value in previous step can be used.

$$\frac{\partial L}{\partial k} = 0 \quad \rightarrow \quad \frac{1}{k} + \frac{\sum_{i=1}^N \ln t_i}{N} - \frac{\sum_{i=1}^N [(\ln t_i) \cdot t_i^k]}{\sum_{i=1}^N t_i^k} = 0 \quad (103)$$

$$\frac{\partial L}{\partial a} = 0 \quad \rightarrow \quad a = \left( \frac{\sum_{i=1}^N t_i^k}{N} \right)^{1/k} \quad (104)$$

$$g(k) = \frac{1}{k} + \frac{T_0}{N} - \frac{T_2(k)}{T_1(k)} \quad g'(k) = -\frac{1}{k^2} - \frac{T_3(k) \cdot T_1(k) - [T_2(k)]^2}{[T_1(k)]^2} \quad (105)$$

$$T_0 = \sum_{i=1}^N \ln t_i \quad T_1(k) = \sum_{i=1}^N t_i^k \quad T_2(k) = \sum_{i=1}^N [\ln t_i \cdot t_i^k] \quad T_3(k) = \sum_{i=1}^N [\ln t_i \cdot \ln t_i \cdot t_i^k] \quad (106)$$

Iterative calculation using following equation shall be carried out untill getting converged value of  $k$ , where  $n$  is a counter for iterative calculation.

$$k_{n+1} = k_n - \frac{g(k_n)}{g'(k_n)} \quad (107)$$

If the value of  $k$  is defined, the value of  $a$  also can be defined using below.

$$a = \left( \frac{\sum_{i=1}^N t_i^k}{N} \right)^{1/k} \quad (108)$$

## 2.9 Exponential distribution

The parameter estimation method in the case of Hydrologic statistic  $x$  following Exponential distribution is shown below.

### (1) Probability Density Function

$$f(x) = \frac{1}{a} \exp\left(-\frac{x-c}{a}\right) \quad (109)$$

$c$  is a location parameter,  $a$  is a scale parameter.

### (2) Cumulative Distribution Function

$$F(x) = 1 - \exp\left(-\frac{x-c}{a}\right) \quad (110)$$

### (3) Quantile $x_p$ for non-exceedance probability $p$

$$p = 1 - \exp\left(-\frac{x-c}{a}\right) \quad \rightarrow \quad x = c - a \ln(1-p) \quad (111)$$

$$x_p = c - a \ln(1-p) \quad (112)$$

### (4) Estimation of parameters (L-moments method)

$$b_0 = \frac{1}{N} \sum_{j=1}^N x_{(j)} \quad b_1 = \frac{1}{N(N-1)} \sum_{j=1}^N (j-1)x_{(j)} \quad (113)$$

where,  $x_{(j)}$  is  $j$ -th value of sample  $x$  in ascending order. It means that  $x_{(1)}$  is the minimum value of sample data and  $x_{(N)}$  is the maximum value of sample data.

The values of  $\lambda_i$  can be calculated using following relationship by deeming  $b_i = \beta_i$ .

$$\lambda_1 = \beta_0 \quad \lambda_2 = 2\beta_1 - \beta_0 \quad (114)$$

Parameters can be estimated using following relationship between L-Moments and parameters.

$$\begin{cases} a = 2\lambda_2 \\ c = \lambda_1 - a \end{cases} \quad (115)$$



## 2.10 Generalized Pareto distribution (GPD)

The parameter estimation method in the case of Hydrologic statistic  $x$  following Generalized pareto distribution is shown below. Exponential distribution is the same as Generalized Pareto distribution with  $k = 0$ .

### (1) Probability Density Function

$$f(x) = \frac{1}{a} \left(1 - k \frac{x-c}{a}\right)^{1/k-1} \quad (k \neq 0) \quad (116)$$

$c$  is a location parameter,  $a$  is a scale parameter,  $k$  is a shape parameter.

### (2) Cumulative Distribution Function

$$F(x) = 1 - \left(1 - k \frac{x-c}{a}\right)^{1/k} \quad (k \neq 0) \quad (117)$$

### (3) Quantile $x_p$ for non-exceedance probability $p$

$$p = 1 - \left(1 - k \frac{x-c}{a}\right)^{1/k} \quad \rightarrow \quad x = c + \frac{a}{k} \{1 - (1-p)^k\} \quad (118)$$

$$x_p = c + \frac{a}{k} \cdot \{1 - (1-p)^k\} \quad (119)$$

### (4) Estimation of parameters (L-moments method)

$$b_0 = \frac{1}{N} \sum_{j=1}^N x_{(j)} \quad b_1 = \frac{1}{N(N-1)} \sum_{j=1}^N (j-1)x_{(j)} \quad b_2 = \frac{1}{N(N-1)(N-2)} \sum_{j=1}^N (j-1)(j-2)x_{(j)} \quad (120)$$

where,  $x_{(j)}$  is  $j$ -th value of sample  $x$  in ascending order. It means that  $x_{(1)}$  is the minimum value of sample data and  $x_{(N)}$  is the maximum value of sample data.

The values of  $\lambda_i$  can be calculated using following relationship by deeming  $b_i = \beta_i$ .

$$\lambda_1 = \beta_0 \quad \lambda_2 = 2\beta_1 - \beta_0 \quad \lambda_3 = 6\beta_2 - 6\beta_1 + \beta_0 \quad (121)$$

Parameters can be estimated using following relationship between L-Moments and parameters.

$$\begin{cases} k = \frac{\lambda_2 - 3\lambda_3}{\lambda_2 + \lambda_3} \\ a = (1+k)(2+k)\lambda_2 \\ c = \lambda_1 - (2+k)\lambda_2 \end{cases} \quad (122)$$

### 3. Confirmation of fitness

#### 3.1 Correlation Coefficient

In order to confirm the fitness of the selected probability distribution model, Q-Q (quantile-quantile) plot is used and the correlation coefficient is written in the graph. In this section, equations to calculate the correlation coefficient is shown.

$$r = \frac{N \sum X_i Y_i - \sum X_i \cdot \sum Y_i}{\sqrt{[N \sum X_i^2 - (\sum X_i)^2] \cdot [N \sum Y_i^2 - (\sum Y_i)^2]}}$$

where,  $r$  : correlation coefficient  
 $N$  : number of a sample  
 $X_i$  : observed data  
 $Y_i$  : calculated data using probability distribution model

#### 3.2 SLSC

$$SLSC = \frac{\sqrt{\frac{1}{N} \cdot \sum_{j=1}^{j=N} (s_j - r_j)^2}}{|r_{0.99} - r_{0.01}|} \quad (123)$$

$s_i$  Normalized variable by parameters  
 $r_i$  Normalized variable by Plotting position formula  
 $r_{0.99}$  Normalized value corresponding to the non-exceedance probability of 99%  
 $r_{0.01}$  Normalized value corresponding to the non-exceedance probability of 1%

Distribution	$S_i$	$r_i$
LN3	$\frac{\ln(x_i - a) - \mu_y}{\sigma_y}$	$qnorm(p_i)$ (%-point of SND)
LP3	$\frac{\ln x_i - c}{a}$	$qgamma(p_i, shape = b, rate = 1)$ (%-point of Gamma Distribution)
Gumbel	$\exp\left(-\frac{x_i - c}{a}\right)$	$-\ln(p_i)$
GEV	$\left(1 - k \frac{x_i - c}{a}\right)^{1/k}$	$-\ln(p_i)$
SQRT-ET	$a \cdot \exp\{\ln(1 + \sqrt{bx_i}) - \sqrt{bx_i}\}$	$-\ln(p_i)$
Weibull	$\left(\frac{x_i - c}{a}\right)^k$	$-\ln(1 - p_i)$
Exponential	$\frac{x_i - c}{a}$	$-\ln(1 - p_i)$
GPD	$-\frac{1}{k} \cdot \ln\left(1 - k \cdot \frac{x_i - c}{a}\right)$	$-\ln(1 - p_i)$

#### 4. Statistical estimation by Jackknife method

The method to obtain the bias-corrected estimate and standard error by Jackknife method is shown in this section. The procedure by Jackknife method is shown below.

- ① Obtain the statistic estimate  $\hat{\theta}$  using observed sample  $X = (x_1, x_2, \dots, x_N)$  with sample size of  $N$ .
- ② Obtain the statistic estimate  $\hat{\theta}_{(i)}$  using sample  $X_i = (x_1, x_2, \dots, x_{i-1}, x_{i+1}, \dots, x_N)$  with sample size of  $N - 1$ . In the dataset  $X_i$ ,  $i$ -th data  $x_i$  is removed.
- ③ By repeating above, dataset  $(\hat{\theta}_{(1)}, \dots, \hat{\theta}_{(N)})$  with the sample size of  $N$  can be obtain. Next, obtain the average of dataset  $(\hat{\theta}_{(1)}, \dots, \hat{\theta}_{(N)})$ .

$$\hat{\theta}_{(\cdot)} = \frac{1}{N} \sum_{i=1}^N \hat{\theta}_{(i)} \quad (124)$$

- ④ Obtain the **bias-corrected jackknife estimate** using following equation.

$$\bar{\theta} = N \cdot \hat{\theta} - (N - 1) \cdot \hat{\theta}_{(\cdot)} \quad (125)$$

where,  $\hat{\theta}$  is the value obtained from step① using original sample  $X = (x_1, x_2, \dots, x_N)$  with sample size of  $N$ , and  $\hat{\theta}_{(\cdot)}$  is the average value obtained from step② and ③ using re-sampled dataset  $X_i = (x_1, x_2, \dots, x_{i-1}, x_{i+1}, \dots, x_N)$  with sample size of  $N - 1$ .

- ⑤ Finally, the **standard error (SE)** of  $\theta$  can be obtained using following equation.

$$(SE) = \sqrt{\frac{N-1}{N} \sum_{i=1}^N (\hat{\theta}_{(i)} - \hat{\theta}_{(\cdot)})^2} \quad (126)$$

#### 5. Statistical estimation by Bootstrap method

The method to obtain the point estimate and confidence interval by Bootstrap method is shown in this section. The procedure by Bootstrap method is shown below.

- ① Set the original observed sample  $X = (x_1, x_2, \dots, x_N)$  with sample size of  $N$ .
- ② Pick randomly  $N$  elements from original sample  $X$  allowing duplicates, and obtain the statistical estimate  $\theta_{*(i)}$  using re-sampled dataset  $X_i$  with sample size of  $N$ .
- ③ Repeat above step and obtain the dataset  $(\theta_{*(1)}, \dots, \theta_{*(B)})$  with sample size of  $B$ . Next, calculate the average of dataset  $(\theta_{*(1)}, \dots, \theta_{*(B)})$ . This becomes **bootstrap point estimate**.

$$\hat{\theta}^* = \frac{1}{B} \sum_{i=1}^B \theta_{*(i)} \quad (127)$$

- ④ It deems that the distribution of a population is the same as the distribution of bootstrap replications  $(\theta_{*(1)}, \dots, \theta_{*(B)})$ . If we set the dataset  $(\theta_{*1}, \dots, \theta_{*B})$  after ranking from bottom to top, **bootstrap confidence interval** at  $1 - \alpha$  level of confidence is  $[\theta_{*\alpha/2 \cdot B}, \theta_{*(1-\alpha/2) \cdot B}]$ . If  $\alpha = 0.05$  and  $B = 1000$ , 95% bootstrap confidence interval becomes  $[\theta_{*25}, \theta_{*975}]$ . This method is called **bootstrap percentile method**.

## 6. Rejection Test

In usual, the very large or small value which seem to be singular in an ordinary sense are to be contained in hydrologic observations. In hydrologic frequency analysis, the rejection test is carried out in order to decide adoption or rejection of the singular variant. In this section, the procedure of rejection test for the singular variant is introduced.

**Subject** Carry out the rejection test for maximum value  $x_\epsilon$  which is contained in a sample with size of  $N$ . (The maximum value  $x_\epsilon$  is the targeted value of the rejection test.)

- ① Calculate the limit singular level  $\epsilon_0$  for the significant level  $\beta_0$  using following equation:

$$\epsilon_0 = 1 - (1 - \beta_0)^{1/N} \quad (\text{normally, } \beta_0=5\%) \quad (128)$$

- ② Select the type of probability distribution model and estimate parameters using the dataset with sample size of  $N - 1$  except the targeted value  $x_\epsilon$  which is seemed singular value.
- ③ Estimate the exceedance probability  $p$  for targeted value  $x_\epsilon$  using adopted probability distribution model.
- ④ Estimate the % point of the standard normal distribution  $u_\epsilon$ , which has equal exceedance probability  $p$  for targeted value  $x_\epsilon$  on adopted probability distribution model. It means to get the value of  $u_\epsilon$  satisfied following condition. Where,  $\Phi$  is the cumulative distribution function of standard normal distribution and  $F$  is the cumulative distribution function adopted for observed sample.

$$\text{Exceedance probability } p = 1 - F(x_\epsilon) = 1 - \Phi(u_\epsilon) \quad (129)$$

- ⑤ If the targeted value belongs to the population to which the sample with size of  $N - 1$  except targeted value belongs, F-value expressed below shall follow the F-distribution with the degrees of freedom of  $(1, M - 1)$ . Note that  $M(= N - 1)$  means sample size except the targeted value.

$$F = \left( \frac{M - 1}{M + 1} \right) \cdot u_\epsilon^2 \quad (130)$$

So, calculate the probability on F-distribution  $2\epsilon$  which satisfies following condition:

$$F_{M-1}^1(2\epsilon) = \left( \frac{M - 1}{M + 1} \right) \cdot u_\epsilon^2 \quad (131)$$

- ⑥ Decide adoption or rejection by comparison between  $\epsilon$  defined by targeted value and  $\epsilon_0$  defined by specified significant level  $\beta_0$ . The detailed criteria is shown below:

$$\left\{ \begin{array}{l} \epsilon \leq \epsilon_0 \quad \text{The exceedance probability for the targeted value is smaller than or equal to} \\ \quad \text{limit singular level.} \\ \quad \text{Therefore, the targeted value shall be rejected. (don't use it in the analysis)} \\ \epsilon > \epsilon_0 \quad \text{The exceedance probability for the targeted value is larger than limit singular level.} \\ \quad \text{Therefore, the targeted value cannot be rejected. (use it in the analysis)} \end{array} \right. \quad (132)$$

(Notice)

### Applying method for the probability distributions except normal

Rejection test method using F-distribution is based of the theorem for only normal distribution. Therefore, steps ②③④ are adopted in order to treat all probability distribution models. In shown above, it is possible to treat all probability distribution models by using the % point of standard normal distribution which has the same exceedance probability for targeted value on selected probability distribution model.

## References

- <http://thesis.ceri.go.jp/center/info/geppou/ceri/0005005050.html>  
(星清：水文統計解析，開発土木研究所月報，No.540，1998年5月 in Japanese character)
- <http://iss.ndl.go.jp/books/R000000004-I4497918-00>  
(星清：現場のための水文統計 (1)，開発土木研究所月報，No.540，1998年5月 in Japanese character)
- <http://iss.ndl.go.jp/books/R000000004-I4522177-00>  
(星清・新目竜一・宮原雅幸：現場のための水文統計 (2)，開発土木研究所月報，No.541，1998年6月 in Japanese character)
- Yoshimi GODA, Masanobu KUDAKA and Hiroyasu KAWAI : Use of L-moments Method for Extreme Statistics of Storm Wave Height, Vol.B2-65, No.1, 2009, 161-165, Journal of JSCE
- Derek A. Roff, ISBN978-4-320-05714-2  
(Derek A. Roff(著)・野間口眞太郎(訳)：生物学のための計算統計学 -最尤法，ブートストラップ法，無作為化法-，共立出版株式会社，2011年3月10日 in japanese character)
- Mutsumi KADOYA : Application of Extreme Value Distribution in Hydrologic Frequency Analysis Part II. Singular Hydrologic Amount and Rejection Test, Citation Bulletins - Disaster Prevention Research Institute, Kyoto University (1964), 66: 33-44, 1964-03-25 (URL <http://hdl.handle.net/2433/123738>)