NATIONAL INSTITUTE OF HYDROLOGY, ROORKEE
WORKSHOP ON
FLOOD FREQUENCY ANALYSIS

TUTORIAL - 2

TESTS OF INDEPENDENCE AND GOODNESS OF FIT

Example 1 :

Test whether annual peak discharge series for Narmada at Garudeshwar (1948-79) given in Tutorial-1 is independent / random or not.

Solution :

Step 1. Calculate lag one serial correlation coefficient by

$$r_1 = \frac{\sum_{t=1}^{N-1} (Y_t - \overline{Y})\,(Y_{t+1} - \overline{Y})}{\sum_{t=1}^{N} (Y_t - \overline{Y})^2} \qquad \text{(T2—1)}$$

where,

r_1 = lag one serial correlation coeff.,

Y_t = annual peak for the t^{th} year.

using eq. T2—1,

$$r_1 = \frac{-41.37 \times 10^7}{684.81 \times 10^7}$$

$$= -0.0604$$

Step 2. Calculate 95% tolerance limits to test the hypothesis of zero auto correlation. 95% tolerance limits are given by

$$\frac{-1 - U_{1+\alpha/2} \cdot (N-K-1)}{N-K} \quad \text{and} \quad \frac{-1 + U_{1-\alpha/2} \cdot (N-K-1)}{N-K}$$

here K = lag = 1 and $U_{1-\alpha/2}$ is normal reduced variate corresponding to (1 – (100—95) / (100 × 2) = 1 – 0.025 = 0.975) prob. of nonexceedance. So limits are

$$\frac{-1-U_{.975} \cdot N-2}{N-1} \quad \text{and} \quad \frac{-1+U_{0.975} \cdot N-2}{N-1}$$

or $\dfrac{-1-1.96 \times 30}{31}$ and $\dfrac{-1+1.96 \times 30}{31}$

or —0.378 and 0.346

Step 3 : Check whether r_1 falls between the limits or not. Since r_1 falls between —0.378 and 0.346, the peak annual discharge series of river Narmada at Garudeshwar can be considered as random.

Example 2 :

Based on D-index test, determine the distribution of peak flood series for Narmada at Garudeshwar. Try only log normal, log Pearson type III and Gumbel EV-1 distributions.

Solution :

$$\text{D-index} = \frac{1}{\bar{X}} \sum_{i=1}^{6} \text{ABS}\,(X_{i,\ \text{observed}} - X_{i,\ \text{computed}})$$

The mean, standard deviation, and coeff. skewness are 29556.9 cumecs, 14864.4 cumecs and 1.052 for the original series and 10.179, 0.488 and 0.1 for log (base e) transformed series.

The highest six observations are 69400, 61350, 58100, 47980, 45630 and 43360 cumecs respectively. In the example highest six observations have been used instead of floods having recurrence interval of 2, 5, 10, 15, 20 and 30 years.

The Weibull plotting position formula has been used to estimate the probability of exceedance.

D indices for log normal, log Pearson type III and Gumbel EV-1,distributions are calculated as below :

Log normal Distribution :

Rank (i)	$X_{i,\ \text{observed}}$ (cumecs)	$P(X \geqslant x)$	Frequency factor (K)	$X_{i,\ \text{computed}}$ (cumecs)	ABS ($X_{i,\ \text{observed}}$ $-X_{i,\ \text{computed}}$)
1	69400	0.0303	1.886	66129	3271
2	61350	0.0606	1·5678	56618	4732
3	58100	0.0909	1.3476	50850	7250
4	47980	0.1212	1,1882	47044	936
5	45630	0.1515	1.0549	44081	1549
6	43360	0.1818	0.9216	41305	2055
					Σ 19793

$$X_{i,\ \text{copmuted}} = e^{(\bar{y} + K.\,\sigma_y)}$$

D-index = 19793/29556.9
= 0.670

Log Pearson Type III Distribution

Rank (i)	$X_{i,\ observed}$ (cumecs)	$P\ (X \geqslant x)$	Frequency factor (K)	$X_{i,\ computed}$ (cumecs)	ABS ($X_{i,\ observed}$ $-X_{i,\ computed}$)
1	69400	0.0303	1.9348	67723	1677
2	61350	0.0606	1.6193	58059	3291
3	58100	0.0909	1.3683	51366	6734
4	47980	0.1212	1.1961	47226	754
5	45630	0.1515	1.0576	44139	1491
6	43360	0.1818	0.9191	41255	2105
					Σ 16052

$$X_{i,\ computed} = e^{(\overline{Y} + K \cdot \sigma_y)}$$

$$\text{D-index} = 16052/29556.9$$
$$= 0.543$$

Gumbel EV—1 Distribution

$$X_{i,\ computed} = \overline{X} - \frac{\sqrt{6}}{\pi}\left(0.5772 + \ln . \ln \frac{T}{T-1}\right) S_X = \overline{X} + K . S_x$$

Rank (i)	$X_{i,\ observed}$ (cumecs)	$P\ (X \geqslant x)$	T	Frequency factor	$X_{i, computed}$ (cumecs)	ABS ($X_{i,\ observed}$ $-X_{i,\ computed}$)
1	69400	0.0303	33.0	2.2642	63213	6187
2	61350	0.0606	16.50	1.7115	55997	5353
3	58190	0.0909	11.00	1.3827	50110	7990
4	47980	0.1212	8.15	1.1455	46584	1396
5	45630	0.1515	6.60	0.8770	42593	3037
6	43360	0.1818	5.50	0.8002	41481	1879
						Σ 25842

$$\text{D-index} = (25842/29556.9)$$
$$= 0.8743$$

D—index is minimum in case of log Pearson Type III distributien. Hence based on D-index test it can be assumed that log Pearson Type III Distribution fits the data well in the upper tail region.

Problem 1 :

Test whether annual peak discharge series for river Narmada at Mortakka (1951—82) is independent/random or not.

Solution :

Computation of lag 1 serial correlation coeff.

Year	Y_t	$Y_t - Y$	$(Y_t - \overline{Y})^2$	$(Y_{t+1} - \overline{Y})$	$(Y_t - \overline{Y})\ (Y_{t+1} - \overline{Y})$
1951					
1952					
1953					
1954					
1955					
1956					
1957					
1958					
1959					
1960					
1961					
1962					
1963					
1964					
1965					
1966					
1967					
1968					
1969					
1970					
1971					
1972					
1973					
1974					
1975					
1976					
1977					
1978					
1979					
1980					
1981					
1982					

$\Sigma Y_t =$ $\qquad$ $\Sigma (Y_t - \overline{Y})^2 =$ $\qquad$ $\Sigma (Y_t - \overline{Y})\ (Y_{t+1} - \overline{Y}) =$

$$r_1 = \frac{\sum_{t=1}^{N-1} (Y_t - \overline{Y})(Y_{t+1} - \overline{Y})}{\sum_{t=1}^{N} (Y_t - \overline{Y})^2}$$

Problem 2 : Based on D-index test, determine the distribution of peak flood series for Narmada at Mortakka. Try normal, log normal, Pearson Type III, log Pearson Type III and Gumbel EV-1 distributions.