LECTURE-3

PROBABILITY DISTRIBUTIONS USED IN FREQUENCY ANALYSIS

OBJECTIVES

The objective of this lecture is to explain various probability distributions used in flood frequency analysis.

3.1 INTRODUCTION

One of the major problems faced in hydrology is the estimation of design flood from fairly short data. If the length of data is more, then the same data can be used to estimate design flood, but the length of data generally available is very less. So the sample data is used to fit frequency distribution which in turn is used to extrapolate trom recorded events to design events either graphically or by estimating the parameters of frequency distribution.

Graphical method is having the advantage of simplicity and visual presentation. But the main disadvantage is that different engineers will fit different curves.

The following continuous distributions are used to fit the annual peak discharge series.

(i) Normal distribution
(ii) Log normal distribution
(iii) Pearson type III distribution
(iv) Exponential distribution
(v) Gamma distribution with two parameters
(vi) Log Pearson type III distribution
(vii) Extreme value distributions

3.2 NORMAL DISTRIBUTION

The normal distribution is one of the most important distributions in statistical hydrology. This is used to fit empirical distributions with symmetrical histograms or with skewness coefficient close to zero. The normal distribution enjoys unique position in the field of statistics due to its role in the central limit theorem. This theorem validates its use as an approximation to other distributions. The central limit theorem states that the distribution of sums of random variables from any distribution tends to a normal distribution as the number of terms in the sum increases.

The probability density function (PDF) of the distribution is given by

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]; -\infty < x < \infty \tag{3.1}$$

where,

μ is the location parameter, and σ is the scale parameter.

The density of normal distribution is continuous $-\infty < x < +\infty$ and tends to zero as x tends to $-\infty$ or $+\infty$. It has a symmetrical bell shape and as a result the mean, mode and median are equal.

The mean, variance, coefficient of skewness and coefficient kurtosis of normally distributed data are

$$E(X) = \mu \tag{3.2}$$

$$Var(X) = \sigma^2 \tag{3.3}$$

$$C_s = 0 \tag{3.4}$$

$$C_k = 3 \tag{3.5}$$

All odd central moments of normal PDF are zero and all even central moments are given by

$$\mu_{2m} = \frac{(2m)!\,\sigma^{2m}}{2^m . m!} \tag{3.6}$$

for example if $m=2$ $\mu_4 = 3\sigma^4$

The cumulative distribution function (CDF) of the normal distribution is given by

$$F(x) = \int_{-\infty}^{x} f(x)\,dx$$

Or

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right) \tag{3.7}$$

In eq. 3.7 if $z = \frac{x-\mu}{\sigma}$; where z is called the reduced variate or standardized variate, then equation reduces to

$$F(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} \exp\left(-\frac{z^2}{2}\right) dz \tag{3.8}$$

The PDF and CDF of normal distribution corresponding to standardized variate z can be calculated from the tables given in Appendix-I and II.

Example 3.1 :

The variate X is normally distributed with $\mu = 10$, and $\sigma = 3.0$. What is the value of x corresponcing to 5 year return period ?

Solution :

$$\text{Return period } T = \frac{1}{\text{Probability of exceedance}}$$

$$= \frac{1}{1\text{-probability of nonexceedance}}$$

$$= \frac{1}{1-F(z)}$$

for $T = 5$, $F(z) = 1 - 1/5 = 0.8$

The value of z corresponding to F (z) = 0.8 is 0.8416 (from Appendix-II). 0.8416 is the z vaiue corresponding to area equal to (0.8 − 0.5 = 0.3).

$$z = \frac{x-\mu}{\sigma}$$

So $x = \mu + \sigma.z$

$= 10 + 3\ (0.8416)$

$= 12.5248$

so the value of x corresponding to 5 years return period is 12.5248.

3.3 LOG NORMAL DISTRIBUTION

Log normal distribution can be applied to a wide variety of hydrologic events especially in the cases in which the corresponding variable has a lower bound, the empirical frequency distribution is not symmetrical and the factors causing those events are independent and multiplicative. Chow (1964) has provided a theoretical justification for the use of the log normal distribution. The causative factors for many hydrologic variables act multiplicatively rather than additively and so the logarithms of the peak flows which are the product of these causative factors follow the log normal distribution.

3.3.1 Two Parameter Lognormal Distribution

If $Y = \log_e (X)$ follows normal distribution. Then X follows log normal disribution Assuming that the mean and variance of Y are μ_y and σ^2_y respectively, the PDF of X can be written as :

$$f(x) = \frac{1}{\sigma_y \sqrt{2\pi}\, x} \exp\left[-\frac{1}{2}\left(\frac{\log_e (x) - \mu_y}{\sigma_y}\right)^2\right] \tag{3.9}$$

In relation to the variable X, μ_y controls the scale, so it is called the scale parameter, while σ_y controls the skewness and hence it may be regarded as a shape parameter.

If log transformed series follows normal distribution then following relationships exist between parameters of $\log_e$ transformed series and original series (Yevjevich, 1972).

$$\mu_y = \tfrac{1}{2} \log_e (\mu^2_x/(1+\eta_x^2)) \tag{3.10}$$

$$\sigma_y^2 = \log_e (1+\eta_x) \tag{3.11}$$

where η_x is the coefficient of variation, $\eta_x = \dfrac{\sigma_x}{\mu_x}$.

The skewness and kurtosis coefficient af the variable X are always positive and are given by

$$C_{s_x} = \eta_x^3 + 3\eta_x \tag{3.12}$$

$$C_{k_x} = \eta_x^8 + 6\eta^6_x + 15\eta^4_x + 16\eta^2_x + 3 \tag{3.13}$$

3.3.2 Three Parameter Lognormal Distribution

If the variable X has a lower bound x_o , different from zero, and the variable $Z = x - x_o$ follows a lognormal distribution with two parameters then X is lognormally distributed with three parameters. The lognormal —3 PDF can be written as

$$f(x) = \frac{1}{\sqrt{2\pi}\,(x - x_o)\,\sigma_y} \exp\left[-\frac{1}{2}\left(\frac{\log_e (x - x_o) - \mu_y}{\sigma_y}\right)^2\right] \tag{3.14}$$

In equation (3.14) μ_y , σ_y and x_o are called the scale (mean of $\log_e (x - x_o)$) the shape (standard deviation of $\log_e (x - x_o)$), and the location parameters, respectively. Parameter x_o is estimated by trial and error.

The lognormal — 3 PDF can be applied to positive or negative valued events provided $x - x_o > 0$, while the lognormal-2 distribution should always be applied to positivs valued events.

The cumulative density function CDF of the lognormal —3 distribution is given by

$$F(x) = \int_{x_o}^{x} \frac{1}{(x-x_o)\sqrt{2\pi}\,\sigma_y} \exp\left(-\frac{1}{2}\left(\frac{\log_e (x - x_o) - \mu_y}{\sigma_y}\right)^2\right) dx \tag{3.15}$$

3.4 PEARSON TYPE III DISTRIBUTION

Pearson type III distribution is a three parameter distribution. This is also known as Gamma distribution with three parameters, The PDF of the distribution is given by

$$f(x) = \frac{(x - x_o)^{\gamma - 1}\, e^{-(x-x_o)/\beta}}{\beta^{\gamma}\,\Gamma(\gamma)} \tag{3.16}$$

$$F(x) = \int_{x_o}^{x} \frac{(x - x_o)^{\gamma - 1}\, e^{-(x-x_o)/\beta}}{\beta^{\gamma}\,\Gamma(\gamma)}\, dx \tag{3.17}$$

where,

x_o = Location Parameter

β = Scale parameter

γ = Shape parameter

The first three statistical estimates of the data are related to the Pearson type III distribution parameters as

$$\mu = x_o + \beta\gamma \tag{3.18}$$

$$\sigma^2 = \beta^2\gamma \tag{3.19}$$

$$c_s = 2/\sqrt{\gamma} \tag{3.20}$$

3.5 EXPONENTIAL DISTRIBUTION

Exponential distribution is a special case of Pearson Type III distribution when $\gamma = 1$. The PDF is given by

$$f(x) = \frac{1}{\beta} e^{(x-x_o)/\beta} \tag{3.21}$$

The CDF is given by

$$F(x) = \int_{x_o}^{x} f(x)\, dx \tag{3.22}$$

$$F(x) = 1 - e^{-(x-x_o)/\beta} \tag{3.23}$$

Using the method of moments the parameters can be obtained using following equations

$$\mu = x_o + \beta \tag{3.24}$$

$$\sigma^2 = \beta^2 \tag{3.25}$$

The coefficient of skewness is equal to 2. If $(x - x_o)/\beta$ is replaced by y then eq. 3.21 and 3.23 can be written as

$$f(y) = e^{-y} \tag{3.26}$$

$$F(y) = 1 - e^{-y} \tag{3.27}$$

In eq. 3.26 and 3.27, y is known as standardized exponential variate or reduced variate.

Example 3.2 :

The variate X has exponential distribution with mean = 20 and standard deviation = 5. What is the value of x corresponding to F (x) = 0.90 i.e. $T = \frac{1}{1 - F(x)} = 10$ years.

Solution :

$$\mu = x_o + \beta = 20$$

$$\sigma^2 = \beta^2$$

or $\quad \beta^2 = 5^2$

so $\quad \beta = 5$

$$x_o = 20 - \beta = 15$$

$$F(y) = 1 - e^{-y}$$

or $\quad 0.9 = 1 - e^{-y}$

or $\quad y = 2.303$

$$y = (x - x_o)/\beta$$

so $\quad x = x_o + \beta \cdot y$

$$x = 15 + 5 \times 2.303$$

$$= 26.516$$

so the value of x corresponding to F (x) = 0.9 is 26.516.

3.6 GAMMA DISTRIBUTION (Two parameter)

This is a special case of Pearson type III distribution when $x_o = 0$. The PDF of Gamma Distribution is given by

$$f(x) = \frac{x^{\gamma-1} e^{-x/\beta}}{\beta^{\gamma} \Gamma(\gamma)} \text{ ; for } \gamma > 0 \tag{3.28}$$

The CDF is given by

$$F(x) = \int_0^x \frac{x^{\gamma-1} e^{-x/\beta}}{\beta^{\gamma} \Gamma(\gamma)} dx \qquad (\ .29)$$

In eq. 3.28 and 3.29 $\Gamma(\gamma)$ is the complete Gamma function while the integral in eq. 3.29 is known as the incomplete Gamma function which can be calculated from table given in Appendix-III, The distribution has no location parameter while β and γ are scale and shape parameters.

The first three statistical estimates of the data are related to the Gamma distribution parameters as

$$\mu = \beta\gamma \tag{3.30}$$

$$\sigma^2 = \beta^2 \gamma \tag{3.31}$$

$$C_s = 2/\sqrt{\gamma} \tag{3.32}$$

The standardized Gamma varite y is given by $y = x/\beta$

3.7 LOG PEARSON TYPE III DISTRIBUTION

The log Pearson Type III distribution has been widely used in hydrology, in particular for fitting the frequency distribution of flood data. The U.S. Water Resources Council recommends the use of the Log Pearson type III distribution as an attempt to promote a uniform and consistent approach for flood frequency studies. As a result, the use of this distribution has become popular in the United States, and has brought the attention of practicising engineers from Federal, State and local government as well as private organisations.

The probability density function of log Pearson type III distribution is given by

$$f(x) = \frac{1}{|\beta|\,\Gamma(\gamma)\cdot x}\left[\frac{\log_e x - y_o}{\beta}\right]^{\gamma-1} \exp\left[-\frac{\log_e x - y_o}{\beta}\right] \tag{3.33}$$

here, y_o is the location parameter, β is the scale parameter and γ is the shape parameter.

If log transformed series of x follows Pearson type III distribution then x series follows log Pearson type III distribution.

3.8 EXTREME VALUE DISTRIBUTIONS

Just as there is a family of Pearson type III distributions, each member being characterized by a value of γ, there is also a family of EV distributions, each member of which is characterized by the value of a parameter denoted by k. The family can be divided into three classes, corresponding to different ranges of k values. The three classes are referred to as Fisher Tippett type 1, type 2 and type 3. They are also known as EV-1, EV-2 and EV-3 distributions. In practice, k value lie in the range—0.6 to + 0.6. For EV-1 distribution k is zero. For EV-2 and EV-3 distributions the values of k are — ve and + ve respectively.

EV-1 and EV-2 distributions are known as Gumbel and Frechet distributions respectively.

3.8.1 General Properties of Extreme Value Distributions

(a) Suppose x has a EV distribution and Z is maximum of it and suppose there are n samples of x values then n, z values will also have EV distribution.

(b) If $y = \log_e x$ follows EV-1 distribution then x series is said to follow EV-2 distribution or log Gumbel distribution.

(c) If x series follows EV-3 distribution then —x series has Weibull distribution.

3.8.2 EV-1 Distribution or Gumbel Distribution

This is a two parameter distribution widely used in meteorology and hydrology.

The PDF is given by

$$f(x) = \frac{1}{\alpha} \exp\left[- \frac{x-u}{\alpha} - e^{-(x-u)/\alpha} \right] \tag{3.34}$$

In the above equation u and α are known as location and shape parameters of the distribution.

The CDF is given by

$$F(x) = e^{-e^{-(x-u)/\alpha}} \tag{3.35}$$

The parameters of the distribution can be estimated from the following equations using method of moments.

$$\mu = u + 0.5772\,\alpha \tag{3.36}$$

$$\sigma^2 = \left(\frac{\pi^2 \alpha^2}{6} \right) \tag{3.37}$$

for Gumbel distribution coeff. of skewness is equal to 1.139.

If in eq. 3.34 and 3.35 $\left(\frac{x-u}{\alpha}\right)$ is replaced by y (the standardized reduced variate) then

$$g(y) = e^{\left(-y-e^{-y}\right)} \tag{3,38}$$

$$G(y) = e^{-e^{-y}} \tag{3.39}$$

or $\quad y = -\ln(-\ln(G(y)))$

$$= -\ln\left(-\ln\left(1-\frac{1}{T}\right)\right) \tag{3.40}$$

Example 3.3 :

The variable X is a EV-1 variate with mean equal to 81 and standard deviation equal to 23. Then what is the return period T corresponding to x=91.

Solution :

$$u + 0.5772\,\alpha = 81$$
$$\pi^2 \alpha^2/6 = 23 \times 23$$

so $\quad \alpha = 17.94$

and $u = 70.65$

$$y = \frac{97-70.65}{17.94}$$

$$= 1.134$$

$$G(y) = e^{-e^{-1.134}}$$

or $1-1/T = e^{-e^{-1.134}}$

or $T=3.567$ years.

So $x=91$ is corresponding to a return period of 3.567 years.

3.8.3 Extreme Value Type II or EV–2 Distribution

This is also known as log Gumbel distribution. The PDF and CDF of the distribution are given by following equations :

$$f(x) = \frac{1}{\alpha}\left[1-k\left(\frac{x-u}{\alpha}\right)\right]^{\frac{1}{k}-1} e\left[1-k\left(\frac{x-u}{\alpha}\right)\right]^{\frac{1}{k}} \tag{3.41}$$

$$F(x) = e^{-(1-k\,(x-u))/\alpha^{1/k}} \tag{3.42}$$

In the above equation $k > 0$, $\alpha > 0$ and $u + \alpha/k \leqslant x_2 \leqslant \infty$

In eq. 3.41 u is location parameter, α is scale parameter and k is shape parameter.

This distribution has a lower bound equal to $u+\alpha/k$. If in eq. 3.41 and 3.42, $1-\frac{x-u}{\alpha}.k$ is replaced by y (the reduced variate) then equations will reduce to

$$g(y) = y^{\frac{1}{k}-1}\ e^{-y^{1/k}}$$

$$G(y) = e^{-y^{1/k}}$$

Example 3.4 :

Variate x has EV–2 distribution with $u=10$, $\alpha = 4$ and $k = -0.1$. What is the event for $T = 5$.

Solution :

$$G(y) = 0.8$$

or $0.8 = e^{-y^{1/k}}$

$$= e^{-y^{-1/0.1}}$$

or $y = 1.162$

$$y = 1 - \frac{x-u}{\alpha} k$$

or $$1.162 = \frac{x-10}{4} (-0.1)$$

or $x = 16.48$

So the event for $T = 5$ is 16.48

3.8.4 EV—3 Distribution

For this distribution the value of k is +ve. This has got a upper bound equal to $u + \alpha/k$. The distribution is used for drought frequency analysis.

References

1. Chow, V.T. (1964), 'Hand Book of Applied Hydrology', McGraw Hill, New York.

2. Haan, C.T. (1977), 'Statistical Methods in Hydrology', Iowa State University Press, Ames, Iowa.

3. Natural Environment Research Council (1975), 'Flood Studies Report', Vol. 1, Institute of Hydrology, Walling ford, U.K.

4. Salas, J.D. (1983), 'Statistics in Water Resources Engineering', lecture notes, Colorado State University, Fort Collins, U.S.A.

5. Yevjevich, V. (1972), 'Probability and Statistics in Hydrology', Water Resources Publication, Fort Collins, Co.

6. Yevjevich, V. (1972), 'Stochastic Processes in Hydrology', Water Resources Publication, Fort Collins, Co.