10.5 PARAMETER ESTIMATION METHODS

A number of methods have been developed to estimate parameters of hydrologic models. Some commonly used methods in hydrology include: (1) method of moments; (2) method of probability weighted moments; (4) L-moments; (5) maximum likelihood estimation; and (6) least squares method. Each of these methods is discussed here.

10.5.1 Method of Moments

This method is very commonly employed to estimate parameters of linear hydrologic models. This method is based on the premise that when the parameters of a probability distribution are estimated correctly, the moments of the probability density function are equal to the corresponding moments of a sample data. Nash (1959) developed the theorem of moments which relates the moments of input, output and impulse response functions of linear hydrologic models.

Let *X* be a continuous variable and *f(x)* its function satisfying some necessary conditions. The r^{th} moment of *f(x)* about an arbitrary point 'a' is denoted as M_r^a (f). The r^{th} moment of the function *f(x)* can be expressed as

$$M_r^a(f) = \int_{-\infty}^{\infty} (x-a)^r f(x)dx \qquad (10.48)$$

Fig. 10.10 shows the definition of various terms used in the above equation.

Consider the special case when r = 0. In this case, eq.(10.48) gives

$$M_0^a = \int_{-\infty}^{\infty} (x-a)^0 f(x)dx = \int_{-\infty}^{\infty} f(x)dx = 1 \qquad (10.49)$$

Thus, the zero-order moment is the area under the curve defined by *f(x)* subject to $-\infty < x < \infty$. For probability distribution, this area is unity. If $r = 1$, eq. (10.48) yields

$$M_1^a = \int_{-\infty}^{\infty} (x-a)^1 f(x)dx = \mu - a \qquad (10.50)$$

where μ is the mean. If the moment is taken around the origin, then $a = 0$, and the first moment gives the mean. When $a = \mu$, the r^{th} moment about the mean is expressed by

$$M_r^\mu = \int_{-\infty}^{\infty} (x-\mu)^r f(x)dx \qquad (10.51)$$

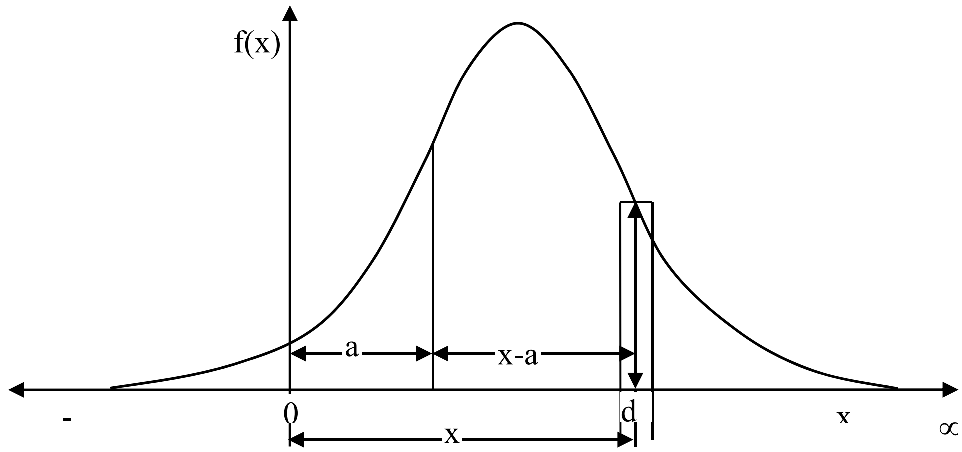

Fig. 10.10 Concept of moment of a function f(x) about an arbitrary point.

For simplicity of notation, we drop the superscript if the moment is taken about the origin 0 and the familiar terminology of the moments can be written as follows:

M_0 =Area
M_1 = Mean
M_2^{μ} = Variance,
M_3^{μ} = Measurement of skewness of the function
M_4^{μ} = Kurtosis,
These terms have already been defined earlier.

10.5.2 Method of Moments for Discrete Systems

For a discrete function, represented as f_j, j = -∞,…, -1,0,1,…, ∞, the r^{th} moment about any arbitrary point can be defined in an analogous manner as for continuous functions. The r^{th} moment about the origin, is defined as

$$M_r = \sum_{m=-\infty}^{\infty} m^r f_m \qquad (10.52)$$

Fig. 10.11 depicts the concept of moment of a discrete function.

Example 10.11: The frequency table of annual flows of Sabarmati River is given Table 10.2. Find the mean and variance of the data by using the method of moments.

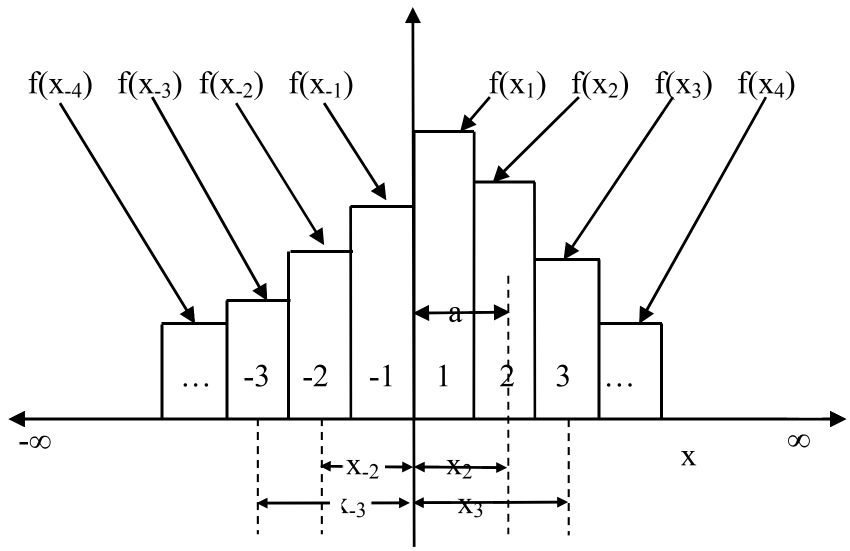

Fig. 10.11 Concept of moment of a discrete function about an arbitrary point 'a'.

Table 10.2 Frequency table of annual flows of Sabarmati River.

Discharge range	Frequency	Discharge range	Frequency
100-200	6	200-300	9
300-400	11	400-500	9
500-600	9	600-700	9
700-800	19	800-900	6
900-1000	6	1000-1100	1
1100-1200	5	1200-1300	2
1300-1400	3	1400-1500	0
1500-1600	2	1600-1700	0
1700-1800	1		

Solution: The sum of all frequencies = (6+11+9+19+ …) = 98. For the first range of discharge (100-200), the mean value is 150; for the second range (200-300), it is 250 and so on.

Hence, the first moment of the data = (150*6 + 250*9 + 350*11 + …+ 1750*1)/98

= 664.29 cumec.

This is the mean (x_m) of the annual flows.

The second moment about the mean gives the variance of the data.

Second $M_2^{\mu} = \sum_{i=1}^{N} \frac{(x_i - x_m)^2 f(x_i)}{N}$ moment

= [$(150\text{-}664)^2$*6 + $(250 – 664)^2$*9 + … + $(1750\text{-}664)^2$*1]/98
= 120000 cumec^2.

Hence, the standard deviation (s) = $(120000)^{0.5}$ =346.61 cumec.

This leads to the coefficient of variation $c_v = s/x_m$ =346.61/664.29 = 0.52.

10.5.3 Method of Maximum Likelihood (MLE)

The maximum likelihood (ML) estimation method is widely accepted as one of the most powerful parameter estimation methods. Asymptotically, the ML parameter estimates are unbiased, have minimum variance, and are normally distributed, while in some cases these properties hold for small samples. The MLE method has been extensively used for estimating parameters of frequency distributions as well as fitting conceptual models.

Let $f(x; a_1, a_2,\ldots a_m)$ be a PDF of the random variable X with parameters a_i, i=1, 2, …, m, to be estimated. For a random sample of data $x_1, x_2, \ldots x_n$, drawn from this probability density function, the joint PDF is defined as

$$f(x_i, x_2, \ldots x_n; a_1, a_2, \ldots a_m) = \mathop{\pi}_{i=1}^{n} f(x_i; a_1, a_2, \ldots a_m) \qquad (10.53)$$

Interpreted conceptually, the probability of obtaining a given value of X, say x_1, is proportional to f(x; a_1, a_2,… a_m). Likewise, the probability of obtaining the random sample x_1, x_2,… x_n from the population of X is proportional to the product of the individual probability densities or the joint PDF. This joint PDF is called the likelihood function, denoted by L.

$$L = \mathop{\pi}_{i=1}^{n} f(x_1; a_1, a_2, \ldots a_m) \qquad (10.54)$$

where the parameters a_i, i=1,2,…m, are unknown.

By maximizing the likelihood that the sample under consideration is the one that would be obtained if n random observations were selected from f(x; a_1, a_2, … a_m), the unknown parameters are determined, and hence the name of the method. The values of parameters so obtained are known as MLE estimators. Since the logarithm of L attains its maximum for the same values of a_i, i = 1, 2, … m, as does L, the MLE function can also be expressed as

$$\ln L = L^* = \ln \mathop{\pi}_{i=1}^{n} f(x_1; a_1, a_2, \ldots a_m) = \sum_{i=1}^{n} \ln f(x_1; a_1, a_2, \ldots a_m)$$

(10.55)

Frequently *ln[L]* is maximized, for it is many times easier to find the maximum of the logarithm of the maximum likelihood function than that of the normal *L*.

The procedure for estimating parameters or determining the point where the MLE function achieves its maximum involves differentiating *L* or *ln L* partially with respect to each parameter and equating each differential to zero. This results in as many equations as the number of unknown parameters. For *m* unknown parameters, we get

$$\frac{\partial L(a_1, a_2, \ldots a_m)}{\partial a_1} = 0$$
$$\frac{\partial L(a_1, a_2, \ldots a_m)}{\partial a_2} = 0 \qquad (10.56)$$
$$\frac{\partial L(a_1, a_2, \ldots a_m)}{\partial a_m} = 0$$

These *m* equations in *m* unknowns are then solved for the *m* unknown parameters.

Example 10.12: Using the method of maximum likelihood, find the parameter α of the exponential distribution for the data of the Sabarmati River in India, given in Example 10.1.

Solution: The probability density function of the one-parameter exponential distribution is given by

$$f_X(x) = \alpha exp(-\alpha x) \qquad (10.57)$$

The likelihood function is given by

$$L(\alpha) = \prod_{i=1}^{n} \alpha \exp(-\alpha x_i) = \alpha^n \exp\left(-\alpha \sum_{i=1}^{n} x_i\right) \qquad (10.58)$$

This can be used to form the log-likelihood function:

$$\ln L(\alpha) = n \ln(\alpha) - \alpha \left(\sum_{i=1}^{n} x_i \right) \qquad (10.59)$$

where *n* is the sample size. Differentiating eq. (10.59) with respect to α:

$$\frac{d \ln L(\alpha)}{d\alpha} = \frac{n}{\alpha} - \sum_{i=1}^{n} x_i = 0$$

This yields

$$\alpha = n \Big/ \left(\sum_{i=1}^{n} x_i \right) = \frac{1}{\bar{x}} \qquad (10.60)$$

In Example 10.1, the mean of the data was found to be 664.29 cumec. Hence, the estimate of α is:

$\alpha = 1/664.29 = 1.51\times10^{-3}$ cumec^{-1}.

10.5.4 Method of Least Squares

The method of least squares (MOLS) is one of the most frequently used parameter estimation methods in hydrology. Natale and Todini (1974) presented constrained MOLS for linear models in hydrology.

Let there be a function Y = f(X; a_1, a_2,… a_m), where a_i, i = 1, 2, … m, are parameters to be estimated. The method of least squares (MOLS) involves estimating parameters by minimizing the sum of squares of all deviations between observed and computed values of *Y*. Mathematically, this sum *D* can be expressed as

$$D = \sum_{i=1}^{n} d_1^2 = \sum_{i=1}^{n} [y_0(i) - y_c(i)]^2 = \sum_{i=1}^{n} [y_0(i) - f(x; a_1, a_2, \ldots a_m)]^2 \qquad (10.61)$$

where $y_0(i)$ is the ith observed value of *Y*, $y_c(i)$ is the ith computed value of *Y*, and $n > m$ is the number of observations. The minimum of *D* in eq. (71) can be obtained by differentiating *D* partially with respect to each parameter and equating each differential to zero, e.g.,

$$\frac{\partial \sum_{i=1}^{n} [y_0(i) - f(x_i; a_1, a_2, \ldots a_m)]^2}{\partial a_1} = 0 \qquad (10.62)$$

The resulting *m* equations, usually called the normal equations, are then solved for estimation of *m* parameters. This method is frequently used to estimate parameters of linear regression model.

10.5.5 Method of L-Moments

Greenwood et al. (1979) introduced the method of probability weighted moments (PWM) and showed its usefulness in deriving explicit expressions for parameters of distributions whose inverse forms X=X (F) can be explicitly defined. They derived relations between parameters and PWMs for Generalized Lambda, Wakeby, Weibull, Gumbel, Logistic and Kappa distributions. However, the probability-weighted moments characterize a distribution but are not meaningful by themselves.

L-moments were developed by Hosking (1986) as functions of PWMs which provide a descriptive summary of the location, scale, and shape of the probability distribution. These moments are analogous to ordinary moments and are expressed as *linear* combinations of order statistics, hence the name. They can also be expressed by linear combinations of probability-weighted moments. Thus, the ordinary moments, the probability weighted moments, and L-moments are related to each other. L-moments are known to have several important advantages over ordinary moments. L-moments have less bias than ordinary moments because they are linear combinations of ranked observations. As an example, variance (second moment) and skewness (third moment) involve squaring and cubing of observations, respectively, which compel them to give greater weight to the observations far from the mean. As a result, they result in substantial bias and variance.

The first L-moment denoted as λ_1 is the arithmetic mean:

$$\lambda_1 = E[X] \qquad (10.63)$$

Let us consider a sample of size n and arrange the data such that $X_{(i|n)}$ is the i^{th} largest observation; clearly $i = n$ will be the largest value. Then, for any distribution, the second L-moment, λ_2, is a description of scale based upon the expected difference between two randomly selected observations:

$$\lambda_2 = (1/2)\ E[X_{(2|1)} - X_{(1|2)}] \qquad (10.64)$$

To compute L-moment measures of skewness three randomly selected observations are used and for kurtosis, we use four randomly selected observations.

$$\lambda_3 = (1/3)\ E[X_{(3|3)} - 2X_{(2|3)} + X_{(1|3)}] \qquad (10.65)$$

$$\lambda_4 = (1/4)\ E[X_{(4|4)} - 3X_{(3|4)} + 3X_{(2|4)} - X_{(1|4)}] \qquad (10.66)$$

Sample L-moment estimates are often computed using (PWMs). The rth PWM is defined (Loucks and Beek, 2005) as:

$$\beta_r = E\{X\ [F(X)]^r\} \qquad (10.67)$$

where $F(X)$ is the cumulative distribution function of X. Recommended (Landwehr et al., 1979; Hosking and Wallis, 1995) unbiased PWM estimators, b_r, of β_r are computed as:

$$
\begin{aligned}
b_0 &= \overline{X} \\
b_1 &= \frac{1}{n(n-1)}\sum_{j=2}^{n}(j-1)X_{(j)} \\
b_2 &= \frac{1}{n(n-1)(n-2)}\sum_{j=3}^{n}(j-1)(j-2)X_{(j)} \\
b_3 &= \frac{1}{n(n-1)(n-2)(n-3)}\sum_{j=4}^{n}(j-1)(j-2)(j-3)X_{(j)}
\end{aligned}
\qquad (10.68)
$$

The general formula for computing estimators b_r of β_r is given by

$$
\begin{aligned}
b_r &= \frac{1}{n}\sum_{j=r+1}^{n}\binom{j-1}{r}X_{(j)}\Big/\binom{n-1}{r} \\
&= \frac{1}{r+1}\sum_{j=r+1}^{n}\binom{j-1}{r}X_{(j)}\Big/\binom{n}{r+1}
\end{aligned}
\qquad (10.69)
$$

for $r = 1, \ldots, n-1$.

L-moments are easily calculated in terms of PWMs using:

$$
\begin{aligned}
\lambda_1 &= \beta_0 \\
\lambda_2 &= 2\beta_1 - \beta 0 \\
\lambda_3 &= 6\beta_2 - 6\beta_1 + \beta_0 \\
\lambda_4 &= 20\beta_3 - 30\beta_2 + 12\beta_1 - \beta_0
\end{aligned}
\qquad (10.70)
$$

As with traditional product moments, measures of the coefficient of variation, skewness and kurtosis of a distribution can be computed with L-moments. Following L-moment ratios are important:

L- coefficient of variation (L-CV) $t_2 = \lambda_2 / \lambda_1$

L- coefficient of skewness (L-sk) $t_3 = \lambda_3 / \lambda_2$

L- coefficient of kurtosis (L-ku) $t_4 = \lambda_4 / \lambda_2$

Example 10.13: Table 10.3 gives annual discharge data of a river for 36 years. Compute sample L-moments and L-moment ratios, L-CV, L-sk, and L-ku.

Table 10.3 Annual discharge data of a river for 36 years

Year	Discharge	Year	Discharge	Year	Discharge	Year	Discharge
1950	400	1959	1390	1968	2291	1977	1499
1951	1100	1960	3300	1969	1340	1978	2598
1952	900	1961	2190	1970	3200	1979	3487
1953	440	1962	935	1971	2200	1980	1234
1954	3000	1963	785	1972	1014	1981	819
1955	2500	1964	501	1973	1790	1982	1210
1956	760	1965	1123	1974	1140	1983	1510
1957	1250	1966	1581	1975	764	1984	1780
1958	1340	1967	959	1976	783	1985	1398

Solution: Equation (10.68) yields estimates of the first three Probability Weighted Moments:

$$b_0 = 1514.19$$
$$b_1 = 889.16$$
$$b_2 = 655.38$$
$$b_3 = 518.64 \qquad (10.71)$$

The sample L-moments can be calculated using the probability weighted moments to obtain:

$$\hat{\lambda}_1 = b_0 = 1514.19$$
$$\hat{\lambda}_2 = 2b_1 - b_0 = 264.12$$
$$\hat{\lambda}_3 = 6b_2 - 6b_1 + b_0 = 111.53 \qquad (10.72)$$
$$\hat{\lambda}_4 = 20b_3 - 30b_2 + 12b_1 - b_0 = -132.82$$

Thus, the sample estimates of the L-coefficient of variation, t_2, and L-coefficient of skewness, t_3, are:

$$t_2 = 264.12/1514.19 = 0.174$$
$$t_3 = 111.53/264.12 = 0.422 \qquad (10.73)$$
$$t_4 = -132.82/264.12 = -0.502$$

10.6 PROBLEMS OF PARAMETER ESTIMATION

The parameters of a distribution function are estimated from the available sample data. But while doing so, errors may arise due to many reasons. The sample data may contain errors, the assumption underlying a particular method of parameter estimation may not hold good, and there may be truncation and round-off errors. All these may result in errors in estimates of parameter. Each estimate of a parameter is a function of sample parameter data which are observations of a random variable. Thus, the estimate value of the parameter itself is a random variable with certain distribution. An estimate obtained from a given set of values can be regarded as an observed value of the random variable. Thus, the goodness of an estimate can be judged from its distribution.

Several questions arise in parameter estimation. How should we employ the available data to obtain the best estimate? What is the best estimate? Are these estimates unique? A number of statistical properties are available by which to address the above questions. These are discussed below.

Bias

Let the estimate of parameter a be a_c denoted by. Estimate a_c will be called an unbiased estimate of a if the expected value of a, denoted $E(a_c) = a$. In general, an estimate will have a certain bias $b(a)$ depending on a so that

$$E(a_c) = a + b(a) \tag{10.74}$$

An unbiased estimate mean $b(a) = 0$. Note that an individual a_c may not be equal to or close to a even if $b(a) = 0$. Unbiasedness simply implies that the average of many independent estimates of a will be equal to a.

The bias in a given quantity is usually measured in dimensionless terms and is often referred to as standardized bias (or BIAS). Thus, BIAS is defined as

$$BIAS = \frac{E(\hat{a}) - a}{a} \tag{10.75}$$

where $\hat{a}$ is an estimate of parameter or quantile of a. In Monte Carlo experimentation, large numbers of samples of different sizes are generated from a given population. For each sample, then, an estimate of a is obtained. If there are, say, 1000 samples of a given size generated then there are 1000 values of parameter a. Thus, E(a) is the average of the 1000 estimates of a for a given sample size and is estimated as

$$E(\hat{a}) = \sum_{i=1}^{n} \hat{a}_i / n \tag{10.76}$$

where n is the number of samples generated or the number of values of the a estimate. The value of a in eq. (10.75) is the true value of a or the value of parameter a of the population.

Efficiency

An estimate a_c of a is said to be efficient if it is unbiased and its variance is at least as small as that of any other unbiased estimate of a. If there are two estimates of a, say a_1 and a_2, then the relative efficiency of a_1 with respect to a_2 is defined as

$$e = \frac{E(a_1 - a)^2}{E(a_2 - a)^2} \leq 1 \tag{10.77}$$

if $E(a_2 - a)^2 > E(a_1 - a)^2$, then $e \leq 1$. An efficient estimate has $e = 1$. If an efficient estimate exists, it may be approximately obtained by use of the MLE or entropy method.

Standard Error

Another dimensionless performance measure frequently used in hydrology is the standard error (SE), defined as

$$SE = \sigma(\hat{a}) / a \tag{10.78}$$

where σ (.) denotes the standard deviation of a and is computed as

$$\sigma(\hat{a}) = \left[\frac{1}{n-1} \sum_{i=1}^{n} \{\hat{a}_i - E(\hat{a}_i)\}^2 \right]^{1/2} \tag{10.79}$$

where the summations are over n estimates $\hat{a}$ of a. In Monte Carlo experiments, referred to as above, for each sample size, a value of SE is obtained. Thus, this measure is similar to the coefficient of variation.

Root Mean Square Error

The root mean square error (RMSE) is one of the most frequently employed performance measures and is defined for parameter a estimate as

$$RMSE = E\left[(\hat{a} - a)^2\right]^{1/2} / a \qquad (10.80)$$

where E[.] is the expectation of [.]. It can be shown that RMSE is related to BIAS and SE as

$$RMSE = \left[\frac{n-1}{n} SE^2 + BIAS^2\right]^{1/2} \qquad (10.81)$$

Relative Mean Error

Another measure of error in assessing the goodness of fit of hydrologic models is the relative mean error (RME) defined as

$$RME = \frac{1}{N}\left(\sum_{i=1}^{N}\left[\frac{Q_0 - Q_c}{Q_0}\right]^2\right)^{0.5} \qquad (10.82)$$

in which N is the sample size, Q is the observed quantity of a given probability and Q_c is the computed quantity of the same probability. Also, used sometimes is the relative absolute error defined as

$$RAE = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{Q_0 - Q_c}{Q_c}\right| \qquad (10.83)$$