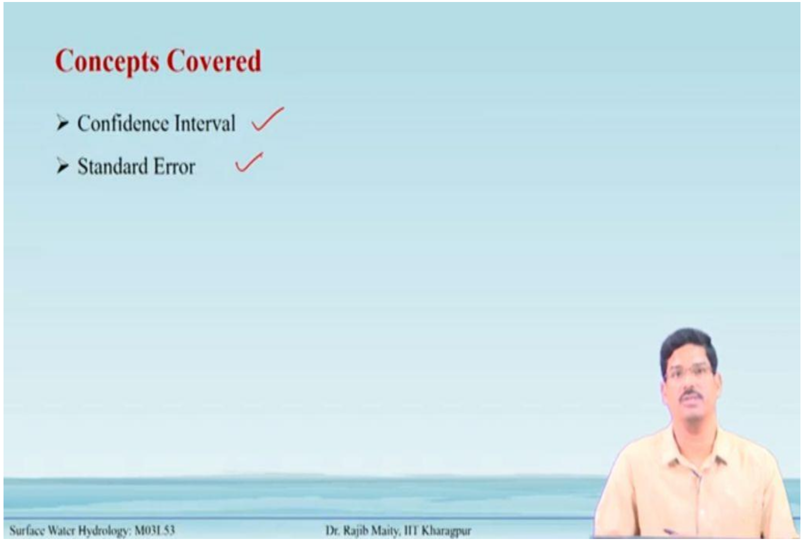
Surface Water Hydrology
Professor Rajib Maity
Department of Civil Engineering
Indian Institute of Technology, Kharagpur
Lecture – 53
Confidence Interval and Standard Error in the Frequency Estimates

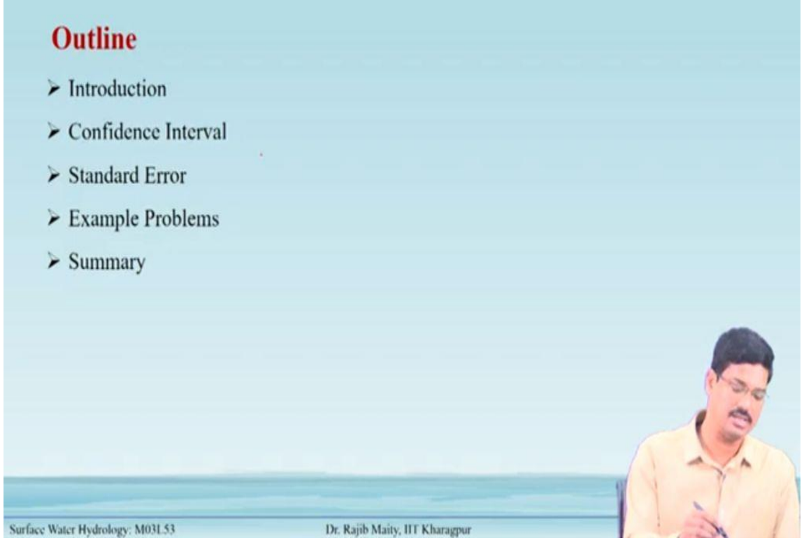
In this today's lecture, we will cover two important concepts that are called Confidence Interval and Standard Error in the Frequent Estimates.

(Refer Slide Time: 00:41)



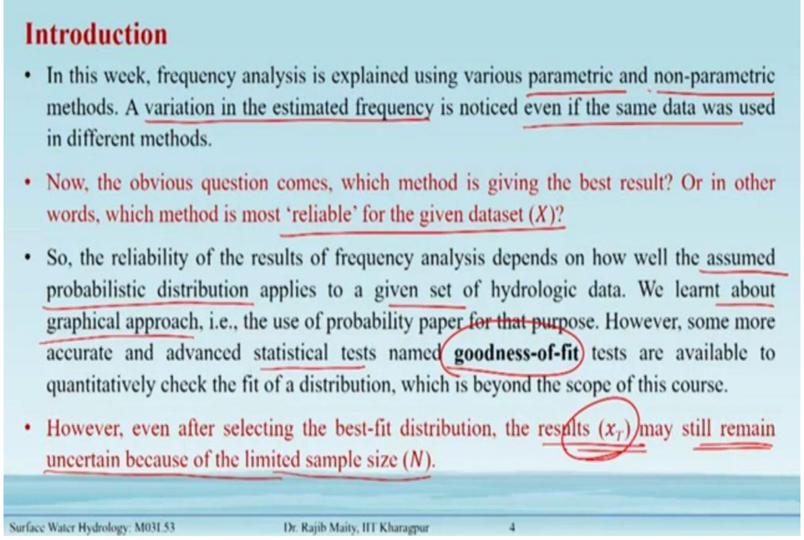
In this lecture we will cover two concepts, the first one is confidence interval and the second one is the standard error, both are with respect to that estimate that we get from frequency analysis.

(Refer Slide Time: 00:53)



The outline goes like this. First, we will give some introduction that why this quantification is necessary, and then the mathematical formulation for the confidence interval and then a standard error will be discussed. And for each of these cases means convince confidence interval and standard error, we will take some example problems, and after that, we will summarize what we learn.

(Refer Slide Time: 01:19)



Introduction

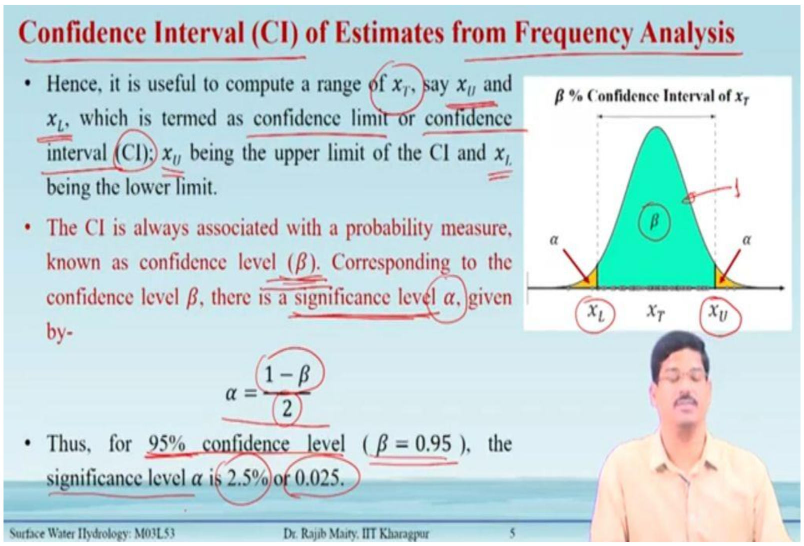
In this week, frequency analysis is explained using various parametric and non-parametric methods. A variation in the estimated frequency is noticed even if the same data was used in different methods.

Now, the obvious question comes, which method is giving the best result? Or in other words, which method is most 'reliable' for the given dataset (X)?

So, the reliability of the results of frequency analysis depends on how well the assumed probabilistic distribution applies to a given set of hydrologic data. We learned about the graphical approach, i.e., the use of probability paper for that purpose. However, some more accurate and advanced statistical tests named goodness-of-fit tests are available to quantitatively check the fit of a distribution, which is beyond the scope of this course.

However, even after selecting the best-fit distribution, the results (x_T) may still remain uncertain because of the limited sample size (N). This is just one example of the reason, there is a limited sample size and other issues also. For example, even if I say that this data follow a normal distribution, then means the question comes how best the fitting is. Is it and perfect normal distribution, which is generally not reliable from the data that we are collecting from different field data. So, the question remains that even if we do all, we take care of all other things, the results that x_T should have some uncertainty associated with it.

(Refer Slide Time: 04:49)



Confidence Interval (CI) of Estimates from Frequency Analysis

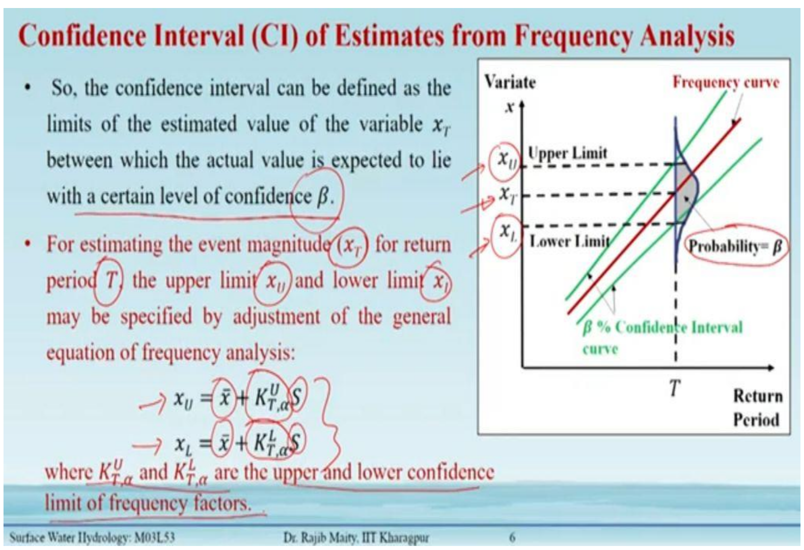
Hence, it is useful to compute a range of x_T, say x_U and x_L, which is termed as confidence limit or confidence interval (CI); x_U being the upper limit of the CI and x_L being the lower limit.

The CI is always associated with a probability measure, known as confidence level (β). Corresponding to the confidence level β, there is a significant level α, given by-

$$\alpha = \frac{1-\beta}{2}$$

Thus, for a 95% confidence level ($\beta = 0.95$), the significance level α is 2.5%, or 0.025.

(Refer Slide Time: 07:53)



Confidence Interval (CI) of Estimates from Frequency Analysis

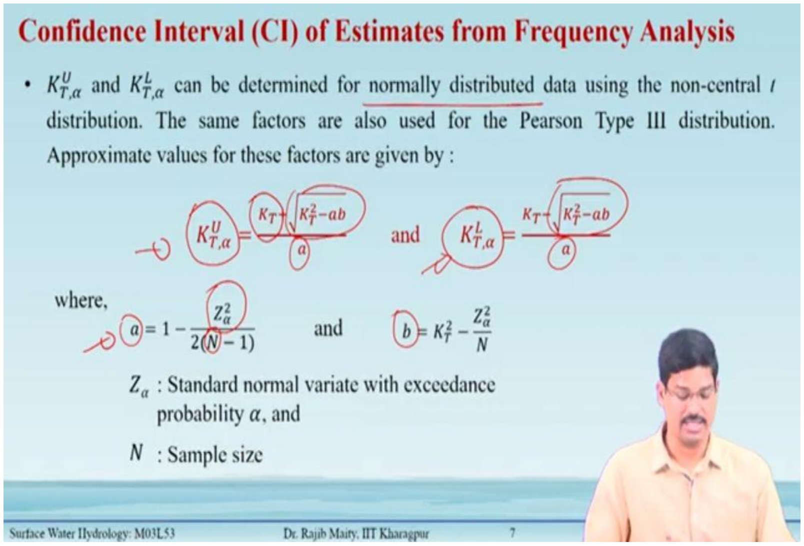
- So, the confidence interval can be defined as the limits of the estimated value of the variable x_T between which the actual value is expected to lie with a certain level of confidence β.
- For estimating the event magnitude (x_T) for return period T, the upper limit x_U and lower limit x_L may be specified by adjustment of the general equation of frequency analysis:

$$x_U = \bar{x} + K^U_{T,\alpha} S$$

$$x_L = \bar{x} + K^L_{T,\alpha} S$$

where $K^U_{T,\alpha}$ and $K^L_{T,\alpha}$ are the upper and lower confidence limit of frequency factors.

(Refer Slide Time: 10:28)



$K^U_{T,\alpha}$ and $K^L_{T,\alpha}$ can be determined for normally distributed data using the non-central t distribution. The same factors are also used for the Pearson Type III distribution. Approximate values for these factors are given by:

$$K^U_{T,\alpha} = \frac{K_T + \sqrt{K_T^2 - ab}}{a} \quad \text{and} \quad K^L_{T,\alpha} = \frac{K_T - \sqrt{K_T^2 - ab}}{a}$$

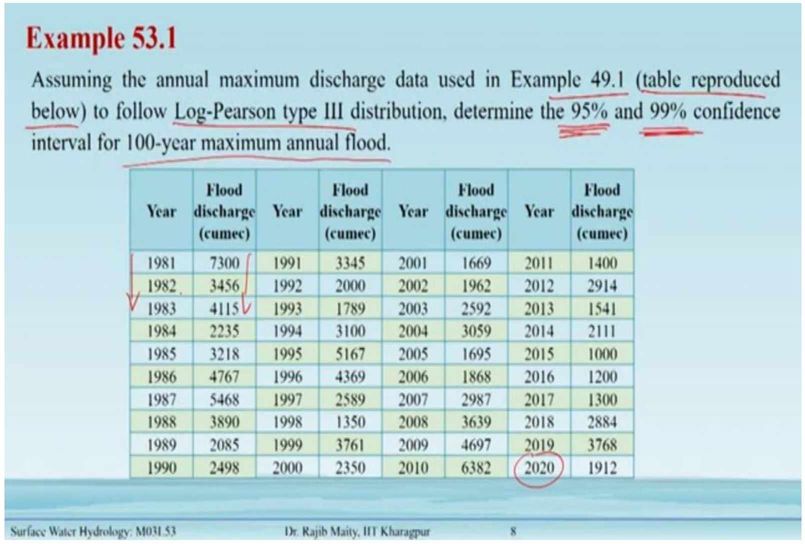
Where,

$a = 1 - \frac{Z_\alpha^2}{2(N-1)}$ and $b = K_T^2 - \frac{Z_\alpha^2}{N}$

Z_α: Standard normal variate with exceedance probability α, and

N: Sample size

(Refer Slide Time: 11:48)

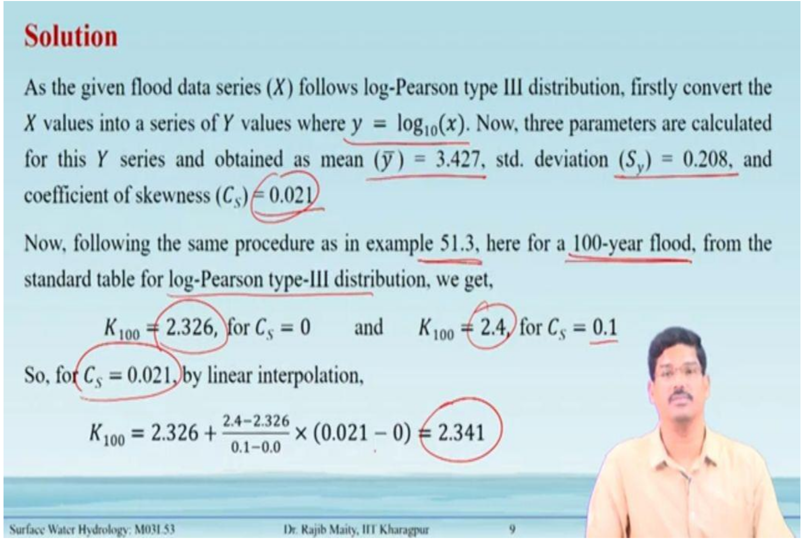


Example 53.1

Assuming the annual maximum discharge data used in Example 49.1 (table reproduced below) to follow Log-Pearson type III distribution, determine the 95% and 99% confidence interval for a 100-year maximum annual flood.

Year	Flood discharge (cumec)	Year	Flood discharge (cumec)	Year	Flood discharge (cumec)	Year	Flood discharge (cumec)
1981	7300	1991	3345	2001	1669	2011	1400
1982	3456	1992	2000	2002	1962	2012	2914
1983	4115	1993	1789	2003	2592	2013	1541
1984	2235	1994	3100	2004	3059	2014	2111
1985	3218	1995	5167	2005	1695	2015	1000
1986	4767	1996	4369	2006	1868	2016	1200
1987	5468	1997	2589	2007	2987	2017	1300
1988	3890	1998	1350	2008	3639	2018	2884
1989	2085	1999	3761	2009	4697	2019	3768
1990	2498	2000	2350	2010	6382	2020	1912

(Refer Slide Time: 12:52)



Solution

As the given flood data series (X) follows the log-Pearson type III distribution, firstly convert the X values into a series of Y values where $y = log_{10}(x)$. Now, three parameters are calculated for this Y series and obtained as the mean ($\bar{y}$) = 3.427, std. deviation (S_y) = 0.208, and coefficient of Skewness (C_S) = 0.021

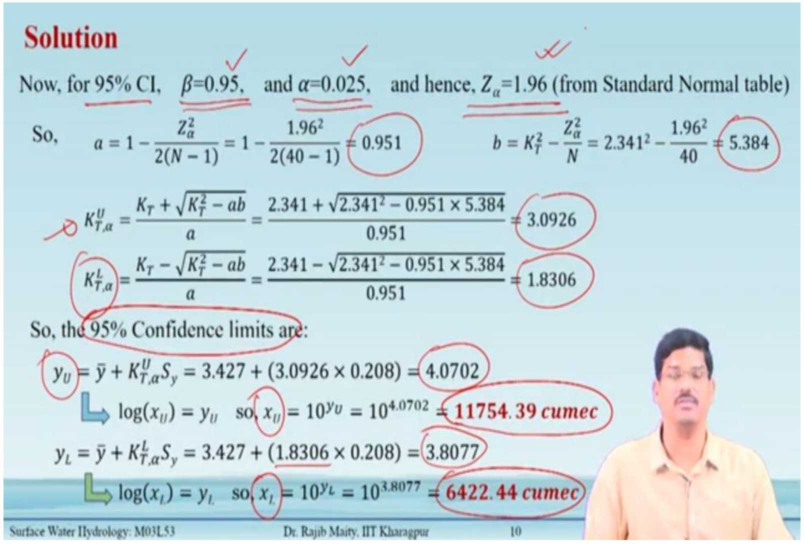
Now, following the same procedure as in example 51.3, here for a 100-year flood, from the standard table for log-Pearson type-III distribution, we get,

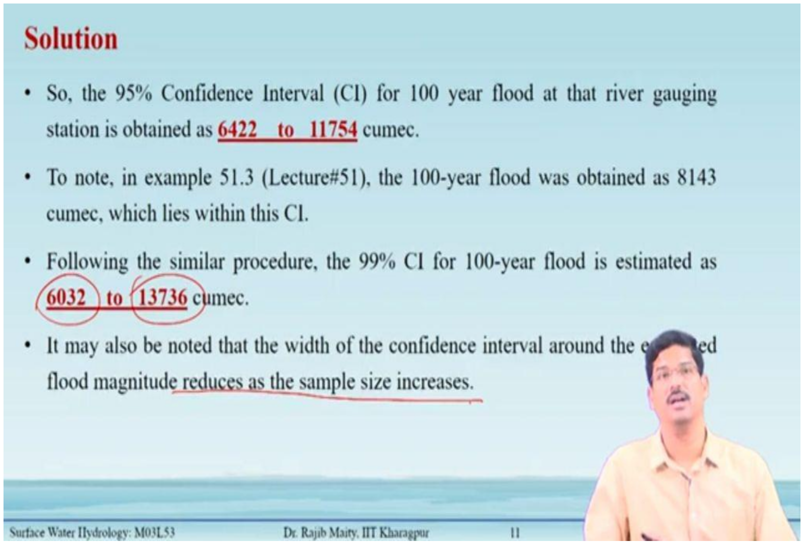
$$K_{100} = 2.326, \text{ for } C_S = 0 \qquad \text{and} \qquad K_{100} = 2.4, \text{ for } C_S = 0.1$$

So, for $C_S =$ 0.021, by linear interpolation,

$$K_{100} = 2.326 + \frac{2.4-2.326}{0.1-0.0} \times (0.021 - 0) = 2.341$$

(Refer Slide Time: 14:11)





Now, for 95% CI, β=0.95, and α=0.025, and hence, Z_{α}=1.96 (from Standard Normal table)

So,

$a = 1 - \frac{Z_{\alpha}^2}{2(N-1)} = 1 - \frac{1.96^2}{2(40-1)} = 0.951$ $\qquad$ $b = K_T^2 - \frac{Z_{\alpha}^2}{N} = 2.341^2 - \frac{1.96^2}{40} = 5.384$

$$K_{T,\alpha}^{U} = \frac{K_T + \sqrt{K_T^2 - ab}}{a} = \frac{2.341 + \sqrt{2.341^2 - 0.951 \times 5.384}}{0.951} = 3.0926$$

$$K_{T,\alpha}^{L} = \frac{K_T - \sqrt{K_T^2 - ab}}{a} = \frac{2.341 - \sqrt{2.341^2 - 0.951 \times 5.384}}{0.951} = 1.8306$$

So, the 95% Confidence limits are:

$$y_U = \bar{y} + K_{T,\alpha}^{U} S_y = 3.427 + (3.0926 \times 0.208) = 4.0702$$

$\log(x_U) = y_U$ so, $x_U = 10^{y_U} = 10^{4.0702} = \mathbf{11754.39}\ \boldsymbol{cumec}$

$$y_L = \bar{y} + K_{T,\alpha}^{L} S_y = 3.427 + (1.8306 \times 0.208) = 3.8077$$

$\log(x_L) = y_L$ so, $x_L = 10^{y_L} = 10^{3.8077} = \mathbf{6422.44}\ \boldsymbol{cumec}$

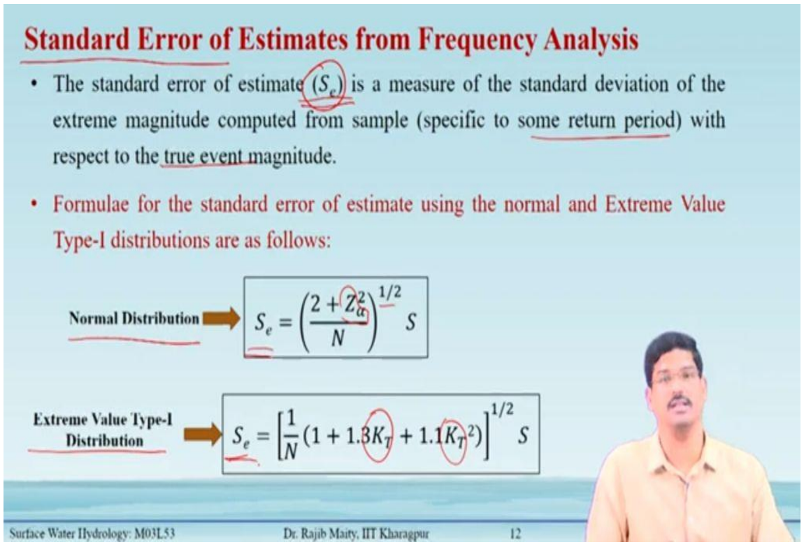
So, the 95% Confidence Interval (CI) for a 100-year flood at that river gauging station is obtained as 6422 to 11754 cumec.

To note, in example 51.3 (Lecture#51), the 100-year flood was obtained as 8143 cumec, which lies within this CI.

Following a similar procedure, the 99% CI for a 100-year flood is estimated as 6032 to 13736 cumec.

It may also be noted that the width of the confidence interval around the estimated flood magnitude reduces as the sample size increases.

(Refer Slide Time: 17:13)



Standard Error of Estimates from Frequency Analysis

The standard error of estimate (Se) is a measure of the standard deviation of the extreme magnitude computed from the sample (specific to some return period) with respect to the true event magnitude.

Formulae for the standard error of estimate using the normal and Extreme Value Type-I distributions are as follows:

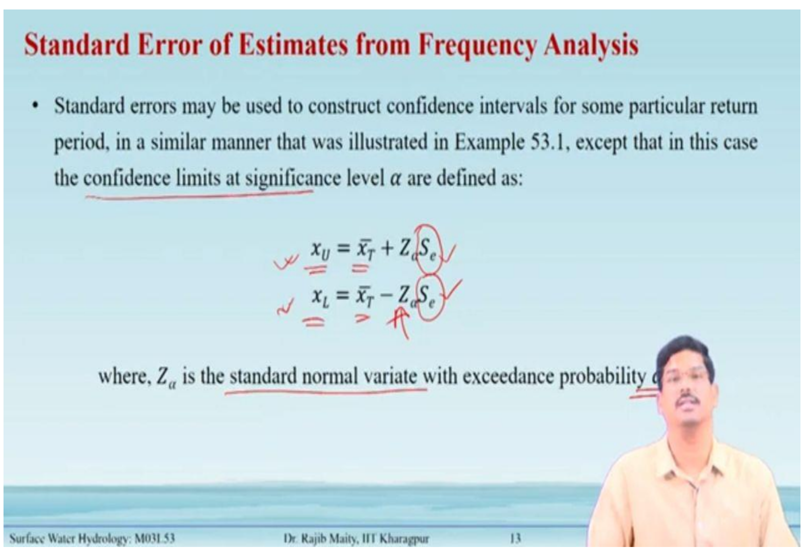
Normal Distribution

$$S_e = \left(\frac{2 + Z_\alpha^2}{N}\right)^{1/2} S$$

Extreme Value Type-I Distribution

$$S_e = \left[\frac{1}{N}(1 + 1.3K_T + 1.1{K_T}^2)\right]^{1/2} S$$

(Refer Slide Time: 18:39)



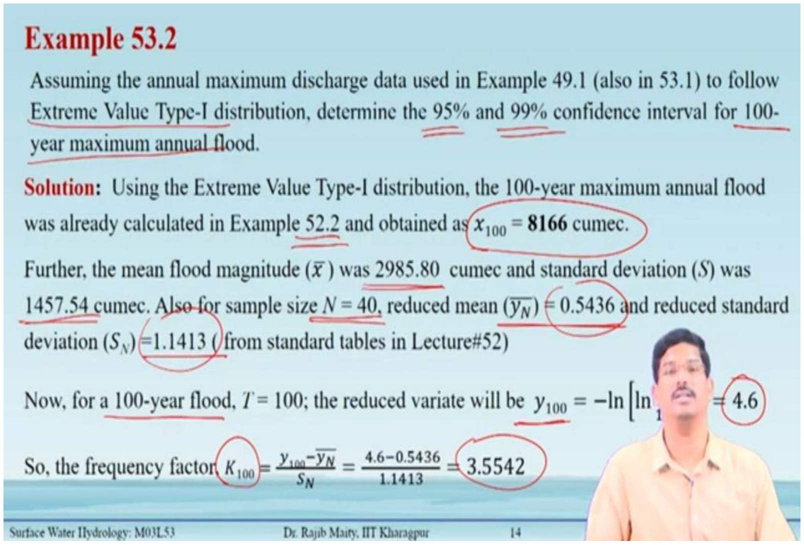
Standard errors may be used to construct confidence intervals for some particular return period, in a similar manner that was illustrated in Example 53.1, except that in this case the confidence limits at significance level α are defined as:

$$x_U = \overline{x_T} + Z_\alpha S_e$$

$$x_L = \overline{x_T} - Z_\alpha S_e$$

where Z_α is the standard normal variate with exceedance probability α.

(Refer Slide Time: 19:34)



Example 53.2

Assuming the annual maximum discharge data used in Example 49.1 (also in 53.1) to follow Extreme Value Type-I distribution, determine the 95% and 99% confidence interval for a 100-year maximum annual flood.

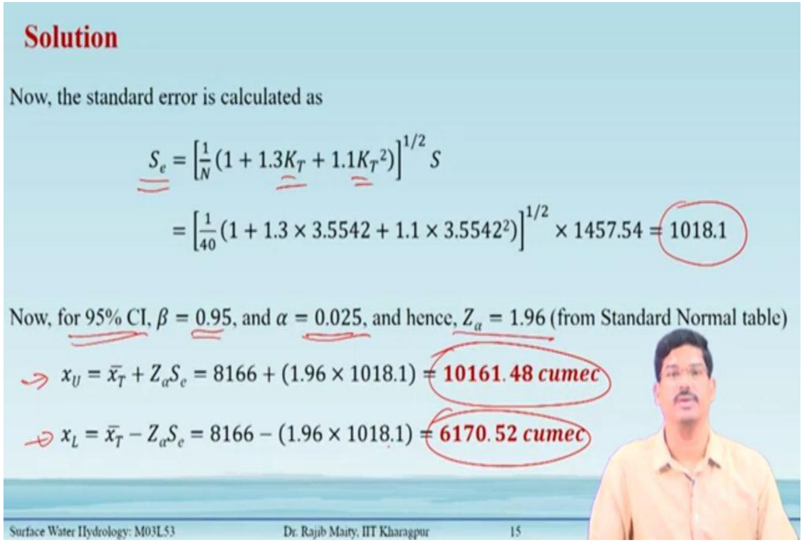
Solution: Using the Extreme Value Type-I distribution, the 100-year maximum annual flood was already calculated in Example 52.2 and obtained as x_{100} = **8166** cumec.

Further, the mean flood magnitude ($\bar{x}$) was 2985.80 cumec and the standard deviation (S) was 1457.54 cumec. Also for sample size N = 40, reduced mean ($\overline{y_N}$) = 0.5436 and reduced standard deviation (S_N) =1.1413 (from standard tables in Lecture#52)

Now, for a 100-year flood, T = 100; the reduced variate will be $y_{100} = -\ln\left[\ln\frac{100}{100-1}\right] = 4.6$

So, the frequency factor, $K_{100} = \frac{y_{100}-\overline{y_N}}{S_N} = \frac{4.6-0.5436}{1.1413} = 3.5542$

(Refer Slide Time: 21:30)



Now, the standard error is calculated as

$$S_e = \left[\frac{1}{N}(1 + 1.3K_T + 1.1{K_T}^2)\right]^{1/2} S$$

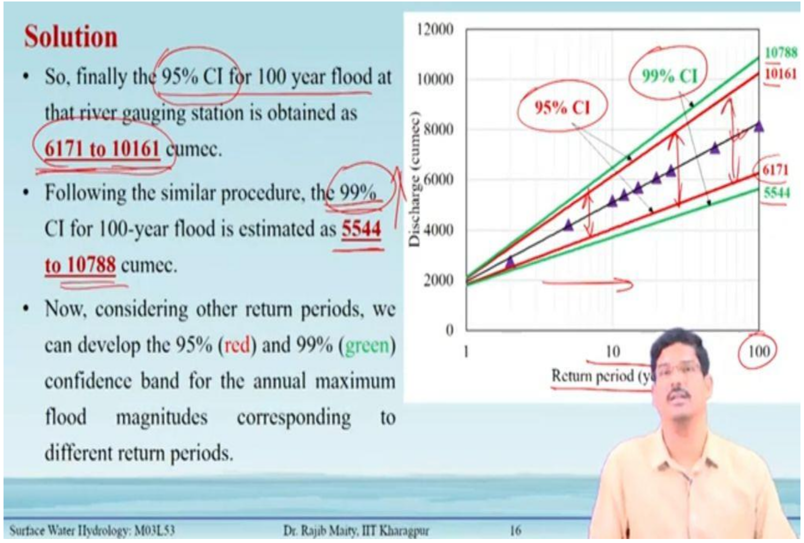
$$= \left[\frac{1}{40}(1 + 1.3 \times 3.5542 + 1.1 \times 3.5542^2)\right]^{1/2} \times 1457.54 = 1018.1$$

Now, for 95% CI, $\beta = 0.95$, and $\alpha = 0.025$, and hence, $Z_\alpha = 1.96$ (from Standard Normal table)

$$x_U = \overline{x_T} + Z_\alpha S_e = 8166 + (1.96 \times 1018.1) = \mathbf{10161.48}\ \boldsymbol{cumec}$$

$$x_L = \overline{x_T} - Z_\alpha S_e = 8166 - (1.96 \times 1018.1) = \mathbf{6170.52}\ \boldsymbol{cumec}$$

(Refer Slide Time: 22:17)



Solution

So, finally, the 95% CI for a 100-year flood at that river gauging station is obtained as 6171 to 10161 cumec.

Following a similar procedure, the 99% CI for a 100-year flood is estimated as 5544 to 10788 cumec.

Now, considering other return periods, we can develop the 95% (red) and 99% (green) confidence band for the annual maximum flood magnitudes corresponding to different return periods.

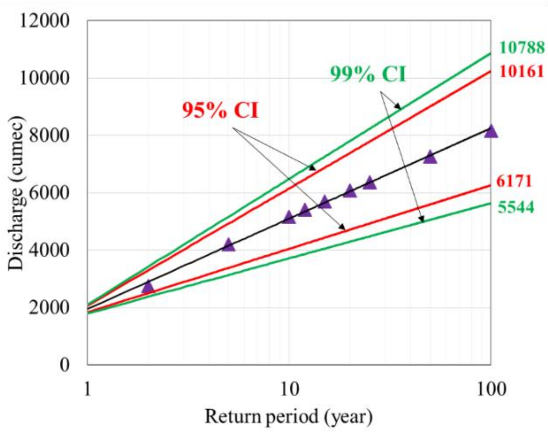
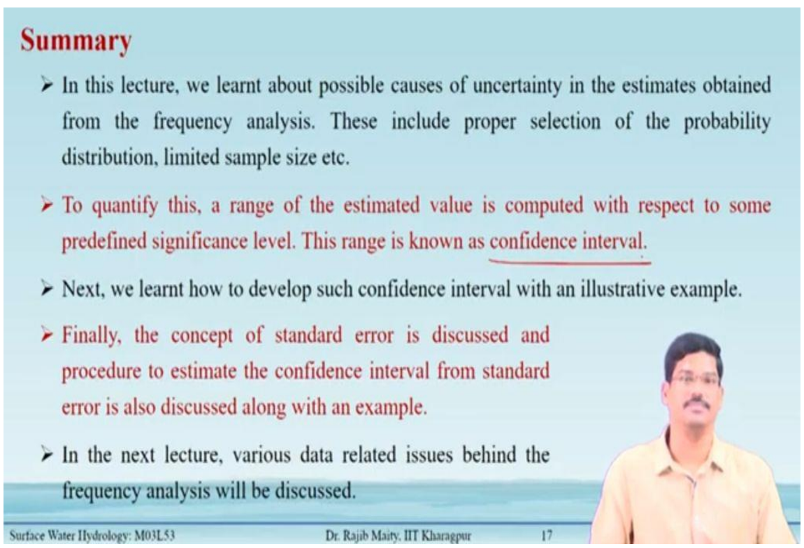


Figure 1 shows the CI for a different return period for example 53.1

(Refer Slide Time: 24:23)



Summary

In summary, we learned the following points from this lecture:

- In this lecture, we learned about possible causes of uncertainty in the estimates obtained from the frequency analysis. These include proper selection of the probability distribution, limited sample size, etc.
- To quantify this, a range of the estimated value is computed with respect to some predefined significance level. This range is known as a confidence interval.

- Next, we learned how to develop such confidence intervals with an illustrative example.
- Finally, the concept of the standard error is discussed and the procedure to estimate the confidence interval from standard error is also discussed along with an example.
- In the next lecture, various data-related issues behind the frequency analysis will be discussed.