INDIAN INSTITUTE OF SCIENCE

STOCHASTIC HYDROLOGY

Lecture -7

Course Instructor : Prof. P. P. MUJUMDAR
Department of Civil Engg., IISc.

Summary of the previous lecture

- Exponential Distribution
- Gamma Distribution
- Extreme Value Distributions
  - Extreme Value Type-I Distribution (Gumbel's Extreme Value Distribution)

Extreme Value Type-I Distribution (Gumbel's Extreme Value Distribution)

$$f(x) = \exp\left\{\mp(x-\beta)/\alpha - \exp\left[\mp(x-\beta)/\alpha\right]\right\}/\alpha$$

$$-\infty < x < \infty;\ -\infty < \beta < \infty;\ \alpha > 0$$

'–' for maximum values and '+' for minimum values

$$\hat{\alpha} = \frac{\sigma}{1.283} \quad ; \quad \hat{\beta} = \mu - 0.45\sigma \quad \text{(maximum)}$$

$$= \mu + 0.45\sigma \quad \text{(minimum)}$$

Y = (X – β)/ α → transformation

•pdf $$f(y) = \exp\left\{\mp y - \exp\left[\mp y\right]\right\} \qquad -\infty < y < \infty$$

•CDF – $$F(y) = \exp\left\{-\exp\left(-y\right)\right\} \quad \text{(maximum)}$$

(Double Exponential Distribution)

$$= 1 - \exp\left\{-\exp\left(y\right)\right\} \quad \text{(minimum)}$$

Example-1
(Gumbel's Extreme Value distribution)

The annual peak flood of a stream exceeds $2000 m^3/s$ with a probability of 0.02 and exceeds $2250 m^3/s$ with a probability of 0.01

1. Obtain the probability that annual peak flood exceeds $2500 m^3/s$

Solution:

The parameters α and β are obtained from the given data as follows

$$F(y) = e^{-e^{-y}} \quad -\infty < y < \infty$$

$$P[X \geq 2000] = 0.02$$

$$P[X \leq 2000] = 0.98$$

$$\text{i.e.,} \quad e^{-e^{-y}} = 0.98$$

Example-1 (contd.)

i.e., $e^{-e^{-y}}$ = 0.98 ; e^{-y}= -ln0.98

y = -ln{-ln(0.98)}

y = 3.902

$$\frac{2000-\beta}{\alpha} = 3.902 \qquad (1)$$

P[X $\geq$ 2250] = 0.01

P[X $\leq$ 2250] = 0.99

exp{-exp(-y)} = 0.99

y = -ln{-ln(0.99)}

y = 4.6

$$\frac{2250-\beta}{\alpha} = 4.6 \qquad (2)$$

Example-1 (contd.)

Solving (1) and (2),

$\alpha = 358$ and $\beta = 603$

Now $P[X \geq 2500] = 1 - P[X \leq 2500]$

$= 1 - \exp\{-\exp(-y)\}$

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

$= (2500-603)/358$

$= 5.299$

$P[X \geq 2500] = 1 - \exp\{-\exp(-5.299)\}$

$= 1 - 0.995$

$= 0.005$

Extreme Value Type-III Distribution

- Referred as Weibull distribution for minimum values
- pdf is given by (for minimum values)

$$f(x) = \alpha x^{\alpha-1} \beta^{-\alpha} \exp\left\{-\left(x/\beta\right)^{\alpha}\right\} \qquad x \geq 0; \alpha, \beta > 0$$

- CDF is given by

$$F(x) = 1 - \exp\left\{-\left(x/\beta\right)^{\alpha}\right\} \qquad x \geq 0; \alpha, \beta > 0$$

- Mean and variance of the distribution are

  $\mu = E[X] = \beta\, \Gamma(1+1/\alpha)$

  $\sigma^2 = Var(X) = \beta^2 \{\Gamma(1+2/\alpha) - \Gamma^2(1+1/\alpha)\}$

Weibull Distribution

- If the lower bound on the parent distribution is not zero, a displacement must be added to the Weibull (type III extreme for minimum) distribution, then pdf is

$$f(x) = \alpha\left(x-\varepsilon\right)^{\alpha-1}\left(\beta-\varepsilon\right)^{-\alpha}\exp\left\{-\left[\left(x-\varepsilon\right)/\left(\beta-\varepsilon\right)\right]^{\alpha}\right\}$$

$$x \geq 0; \alpha, \beta > 0$$

- known as 3-parameter Weibull distribution
- CDF is

$$F(x) = 1-\exp\left\{-\left[\left(x-\varepsilon\right)/\left(\beta-\varepsilon\right)\right]^{\alpha}\right\} \qquad x \geq 0; \alpha, \beta > 0$$

Weibull Distribution

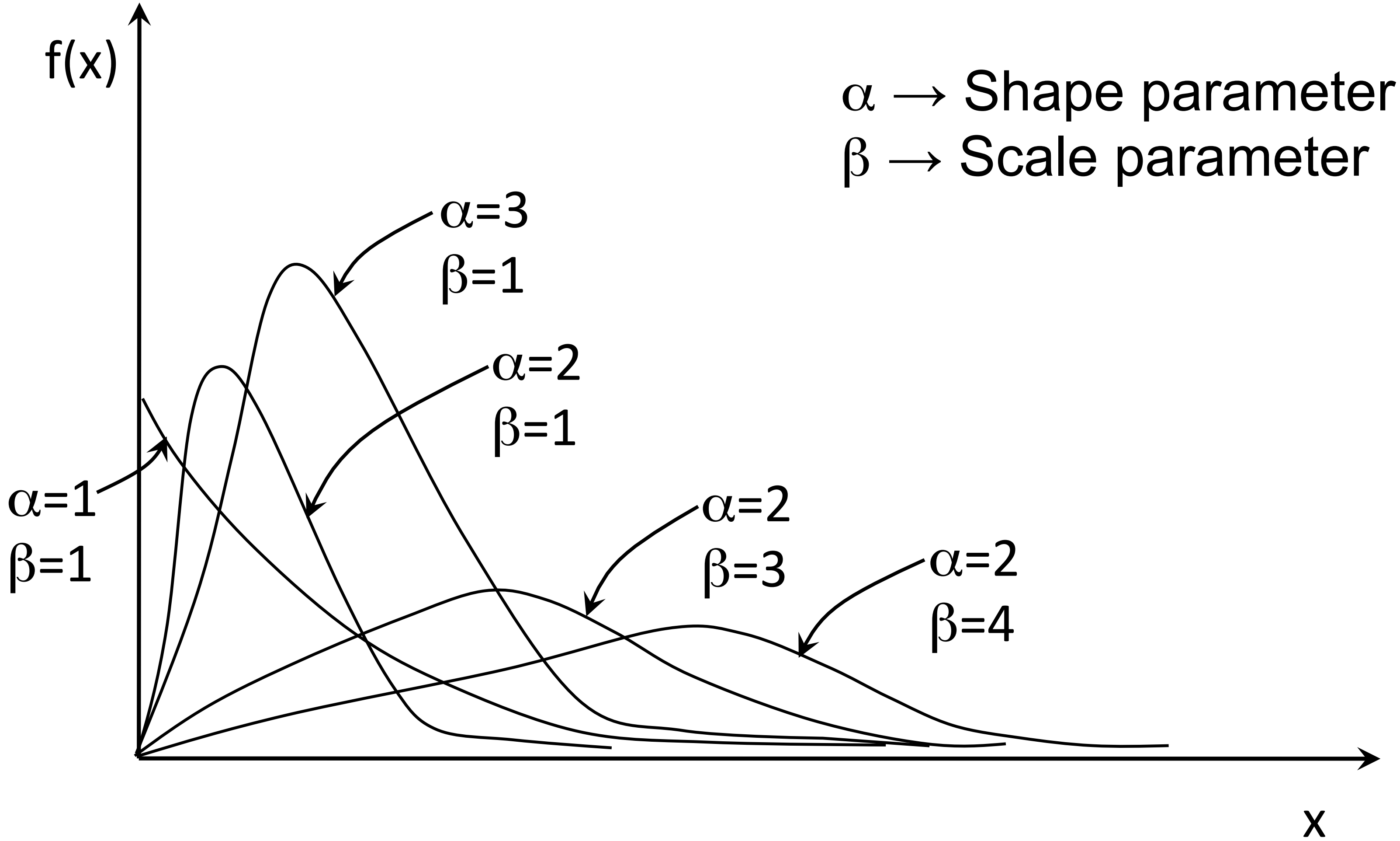
- $Y = \{(X - \varepsilon)/ (\beta - \varepsilon)\}^{\alpha} \rightarrow$ transformation

- Mean and variance of the 3-parameter Weibull distribution are

$$\mu = E[X] = \varepsilon + (\beta - \varepsilon)\, \Gamma(1+1/\alpha)$$

$$\sigma^2 = Var(X) = (\beta - \varepsilon)^2 \{\Gamma(1+2/\alpha) - \Gamma^2(1+1/\alpha)\}$$

Weibull Distribution



Example-2

Obtain P[X $\leq$ 0.1] using Weibul's distribution for a sample X:{0.2, 0.3, 0.45, 0.05, 0.6, 0.07, 0.02, 0.65, 0.15, 0.01}

Sample mean $\bar{x}$ = 0.25

Sample variance s^2 = 0.05782

μ = E[X] = $\beta\ \Gamma(1+1/\alpha)$

$\Gamma(1+\eta) = \eta\sqrt{\eta}$

$$\mu = \beta \times \frac{1}{\alpha}\sqrt{\frac{1}{\alpha}} = \frac{\beta}{\alpha\sqrt{\alpha}}$$

$$\alpha\sqrt{\alpha} = \frac{\beta}{\mu}$$

Example-2 (contd.)

$\sigma^2 = Var(X) = \beta^2 \{\Gamma(1+2/\alpha) - \Gamma^2(1+1/\alpha)\}$

$$\sigma^2 = \beta^2 \left[\frac{2}{\alpha}\sqrt{\frac{2}{\alpha}} - \frac{\mu^2}{\beta^2}\right]$$

$\mu = \beta\, \Gamma(1+1/\alpha)$

$\mu /\beta = \Gamma(1+1/\alpha)$

$$\sigma^2 = 2\sqrt{2}\,\mu\beta - \mu^2$$

$$\beta = \frac{\sigma^2 + \mu^2}{2\sqrt{2}\,\mu} = \frac{0.05782 + 0.25^2}{2\sqrt{2}\times 0.25}$$

Substituting the sample moment values

$$\beta = 0.1702$$

$$\alpha\sqrt{\alpha} = \frac{\beta}{\mu} = \frac{0.1702}{0.25} = 0.68063$$

$$\alpha = 0.774$$

Example-2 (contd.)

$P[X \leq 0.1] = F(0.1)$

$$F(x) = 1 - \exp\left\{-\left(x/\beta\right)^{\alpha}\right\}$$

$$F(0.1) = 1 - \exp\left\{-\left(0.1/0.1702\right)^{0.774}\right\}$$

$$= 0.4845$$

$P[X \leq 0.1] = 0.4845$

PARAMETER ESTIMATION

Parameter Estimation

- f(x) : pdf and F(x) : CDF
- In general, f(x) and F(x) are also functions of parameters

$f(x; \theta_1; \theta_2 \ldots\ldots\ldots\ldots \theta_m)$ or $F(x; \theta_1; \theta_2 \ldots\ldots\ldots\ldots \theta_m)$

- A random sample $x_1, x_2, \ldots\ldots x_n$ is available

  $\hat{\theta}_i$: Estimate of θ_i : a function of the sample
- $\hat{\theta}_i$: random variable since it is a function of the random sample.
- Two important properties of the estimators
  - Unbiasedness
  - Consistency

Parameter Estimation

Unbiased estimate:

- An estimate $\hat{\theta}$ of a parameter θ is said to be unbiased if $E(\hat{\theta}) = \theta$.
- The bias is given by $E(\hat{\theta}) - \theta$
- An estimator is unbiased does not guarantee that an individual $\hat{\theta}$ is equal to θ or even close to θ
- The average of many independent estimates of θ will be equal to θ.

Parameter Estimation

Consistent estimate:

- An estimate $\hat{\theta}$ of a parameter θ is said to be consistent if the probability that $\hat{\theta}$ differs from θ by more than an arbitrary constant ε approaches zero as the sample size approaches infinity.

$$\underset{n\to\infty}{P}\left[\left|\hat{\theta}-\theta\right| \geq \varepsilon\right] \longrightarrow 0$$

Parameter Estimation

Methods of estimating parameters from samples of data:

- Method of matching points:
- Method of moments
- Method of maximum likelihood

Method of matching points

- Not a commonly used method
- Can produce reasonable first approximations to the parameters
- Use the data set to obtain probabilities and estimate the parameters
- Simple and approximate method

Example-3

A data set is assumed to follow exponential distribution

$$f(x) = ce^{-cx} \qquad x > 0$$

In the data set, 80% of the values are less than 1.5

Estimate the parameter ‘c’

$P[X < 5.0] = 0.8$

$1 - e^{-1.5c} = 0.8$

$e^{-1.5c} = 0.2$

$-1.5c = \ln(0.2)$

$1.5c = 1.61$

$c = 1.073$

Method of Moments (MoM)

- One of the most common used methods for estimating the parameters
- Equate the first ‘m’ moments of the population to the sample estimates of the first ‘m’ moments
- Results in ‘m’ equations; solve to get the ‘m’ unknown parameters of the distribution.

Example-4

Obtain the parameter ‘λ’ using method of moments for the pdf $f(x) = \lambda e^{-\lambda x} \qquad x > 0$

The first moment is $\mu = \int_0^\infty x \cdot \lambda e^{-\lambda x}\, dx$

$$= \lambda \left\{ \left[-\frac{xe^{-\lambda x}}{\lambda} \right]_0^\infty - \int_0^\infty -\frac{e^{-\lambda x}}{\lambda} dx \right\}$$

$$= \lambda \left[-\frac{xe^{-\lambda x}}{\lambda} - \frac{e^{-\lambda x}}{\lambda^2} \right]_0^\infty$$

$$= \left[-e^{-\lambda x} \left(x + \frac{1}{\lambda} \right) \right]_0^\infty = \frac{1}{\lambda}$$

Example-4(contd.)

The first moment of the sample is $\bar{x}$

Therefore, $\bar{x} = \frac{1}{\lambda}$

$$\hat{\lambda} = \frac{1}{\bar{x}}$$

Example-5

Obtain the estimate 'θ' using method of moments for pdf

$$f(x) = \theta \cdot \sin^2 x \qquad\qquad 0 \le x \le \pi$$

One parameter ' θ'

The first moment is $\mu = \int_0^{\pi} x \cdot \theta \sin^2 x \, dx$

$$\bar{x} = \int_0^{\pi} x \cdot \theta \sin^2 x \, dx$$

$$\bar{x} = \theta \left[\frac{\sin x^2}{4} - \frac{x \sin 2x}{4} + \frac{x^2}{4} \right]_0^{\pi}$$

Example-5(contd.)

$$\overline{x} = \theta\left(\frac{\pi^2}{4}\right)$$

Therefore the estimate is

$$\hat{\theta} = \frac{4\overline{x}}{\pi^2}$$

Method of Maximum Likelihood

- Consider the sample of 'n' random observations x_1, x_2 ……… x_n
- Joint pdf is $f(x_1, x_2 \ldots x_n; \theta_1; \theta_2 \ldots \theta_m)$
- The Likelihood function is formulated as

$L = f(x_1;\theta_1;\theta_2 \ldots\theta_m) \times f(x_2;\theta_1;\theta_2 \ldots\theta_m) \ldots \times f(x_n;\theta_1;\theta_2 \ldots \theta_m)$

$$= \prod_{i=1}^{n} f\left(x_i, \theta_1, \ldots\ldots..\theta_m\right)$$

Method of Maximum Likelihood

- We are interested in those values of θ_1, θ_2 ………… θ_m that maximize the Likelihood function

$$\frac{\partial L}{\partial \theta_i} = 0 \quad \forall\, i$$

- Solving the 'm' equations, the 'm' parameters are estimated
- Maximum Likelihood (ML) estimators are not unbiased
- It may be shown that they are asymptotically unbiased
- ML estimates are consistent.
- MoM and ML do not always produce the same estimators for parameters.
- ML is generally preferred over MoM.

Example-6

Obtain the parameter ‘λ’ using method of maximum likelihood for the pdf

$$f(x) = \lambda e^{-\lambda x} \qquad\qquad x > 0$$

$$L(\lambda) = \lambda e^{-\lambda x_1} \times \lambda e^{-\lambda x_2} \ldots\ldots\ldots\ldots \lambda e^{-\lambda x_n}$$

$$= \lambda^n e^{-\lambda \sum x_i}$$

$$\ln L(\lambda) = n \ln \lambda - \lambda \sum x_i$$

$$\frac{\partial \ln L(\lambda)}{\partial \lambda} = 0$$

$$\frac{n}{\lambda} - \sum x_i = 0$$

Example-6 (contd.)

$$-\lambda \sum x_i + n = 0$$

$$\lambda \sum x_i = n$$

$$\frac{\sum x_i}{n} = \frac{1}{\lambda}$$

$$\bar{x} = \frac{1}{\lambda}$$

$$\hat{\lambda} = \frac{1}{\bar{x}}$$

Example-7

Obtain the maximum likelihood estimates of the parameter 'β' in the pdf

$$f(x) = 2\beta\sqrt{\frac{\beta}{\pi}}\, x^2\, e^{-\beta x^2} \qquad -\infty < x < \infty$$

$$L(\beta) = 2\beta\sqrt{\frac{\beta}{\pi}}\, x_1^2\, e^{-\beta x_1^2} \times 2\beta\sqrt{\frac{\beta}{\pi}}\, x_2^2\, e^{-\beta x_2^2} \ldots\ldots\ldots\ldots$$

$$= 2^n \beta^n \left(\frac{\beta}{\pi}\right)^{n/2} \left(\prod_{i=1}^{n} x_i^2\right) e^{-\sum_{i=1}^{n} \beta x_i^2}$$

$$= 2^n \beta^{(n+n/2)} \pi^{-n/2} \left(\prod_{i=1}^{n} x_i^2\right) e^{-\sum_{i=1}^{n} \beta x_i^2}$$

Example-7 (contd.)

$$\ln L(\beta) = n\ln 2 + (n + n/2)\ln\beta - \frac{n}{2}\ln\pi + \ln\left(\prod_{i=1}^{n} x_i^2\right) - \beta\sum_{i=1}^{n} x_i^2$$

$$\frac{\partial \ln L(\beta)}{\partial \beta} = 0$$

$$(n + n/2)\frac{1}{\beta} - \sum_{i=1}^{n} x_i^2 = 0$$

$$\frac{3n}{2} = \sum_{i=1}^{n} x_i^2 \times \beta$$

$$\hat{\beta} = \frac{3n}{2\sum_{i=1}^{n} x_i^2}$$

Chebyshev Inequality

- Certain general statements may be made on the rvs without placing restrictions on their distributions.
- Chebyshev inequality states that a single observation selected at random from any probability distribution will deviate more than $k\sigma$ from mean μ with a probability less than or equal to $1/k^2$.

$$P\left[\left|X-\mu\right| \geq k\sigma\right] \leq \frac{1}{k^2}$$

(places an upper bound on the probability.)

Example-8

The mean annual stream flow of a river is 135 Mm^3 and standard deviation is 23.8 Mm^3.What is the maximum probability that the flow in a year will deviate more than 45 Mm^3 from the mean.

Applying Chebyshev inequality, $P\left[\left|X-\mu\right| \geq k\sigma\right] \leq \frac{1}{k^2}$

$$k\sigma = 45$$

$$k \times 23.8 = 45$$

$$k = 1.891$$

$$P\left[\left|X-\mu\right| \geq 45\right] = P\left[\left|X-\mu\right| \geq 1.891\sigma\right] \leq \frac{1}{k^2}$$

$$\leq 1/1.891^2$$

$$\leq 0.28$$