INDIAN INSTITUTE OF SCIENCE

STOCHASTIC HYDROLOGY

Lecture -2

Course Instructor : Prof. P. P. MUJUMDAR

Department of Civil Engg., IISc.

Summary of the previous lecture

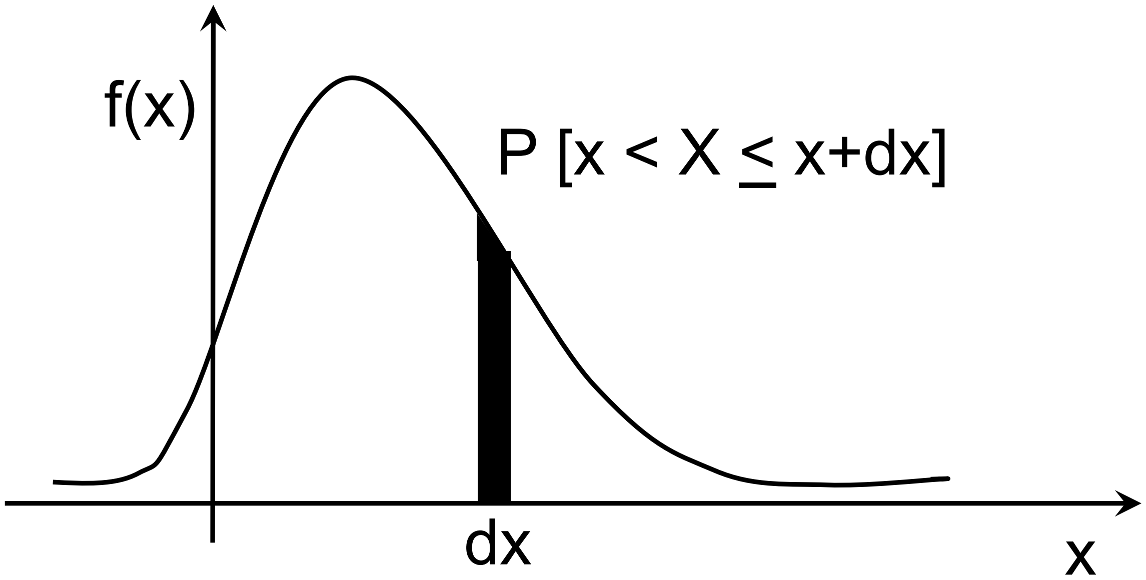
- Concept of a random variable
- Discrete and continuous random variables
- Probability mass function, density function and cumulative distribution functions

Corrections in the slides of Lecture-1

Slide No.	Title of slide	Original	Correction
23	Continuous RVs	$f(x) = \lim_{dx\to\infty} \frac{\mathrm{P}\left[x < X \le x + dx\right]}{dx}$	$f(x) = \lim_{dx\to 0} \frac{\mathrm{P}\left[x < X \le x + dx\right]}{dx}$
32	Mixed distributions	$\int_{-\infty}^{d} f_1(x) + P\left[x = d\right] +$ $\int_{d}^{\infty} f_2(x) = 1.0$	$\int_{-\infty}^{d} f_1(x)dx + P\left[x = d\right] +$ $\int_{d}^{\infty} f_2(x)dx = 1.0$
39		P[5 $\le$ x $\le$ 3]	P[5 $\le$ X $\le$ 3]

Corrections in Lecture-1

Slide No.23-Continuous RVs



$$f(x) = \lim_{dx \to \infty} \frac{P[x < X \leq x + dx]}{dx} \quad \text{where} \int_{-\infty}^{\infty} f(x)dx = 1$$

CORRECTED AS

$$f(x) = \lim_{dx \to 0} \frac{P[x < X \leq x + dx]}{dx} \quad \text{where} \int_{-\infty}^{\infty} f(x)dx = 1$$

Bivariate Distributions

- In many situations, we would be interested in simultaneous behavior of two or more random variables. e.g., in hydrology, we may be interested in the joint behavior of
  - Rainfall – Runoff
  - Rainfall – Recharge
  - Rainfall intensity- Peak flood discharge
  - Temperature – Evaporation
  - Soil permeability – GW yield
  - Flow rates on two streams

Bi-variate Distributions

- We denote (X,Y) as a two-dimensional random variable (or a two dimensional random vector).
- X and Y both discrete : two dimensional discrete r.v
- X and Y both continuous : two dimensional continuous r.v.
- It is possible that one of the rvs of (X, Y), say, X, is discrete while the other is continuous. In this course, however, we deal only with cases in which both X & Y are either discrete or continuous.

Probability distribution of (X, Y)

- We define the ***joint probability mass function*** of a two dimensional discrete r.v., (X,Y) as,

  $p(x_i, y_j) = P[X=x_i, Y=y_j]$

  By this we imply, $P[X = x_i, \text{AND } Y = y_j]$

- $p(x_i, y_j) \geq 0$
- $$\sum_{j=1}^{\infty} \sum_{i=1}^{\infty} p(x_i, y_j) = 1$$

Probability Distribution of (X, Y)

- Similar to the CDF for one dimensional random variables, we define the ***joint CDF*** of the two dimensional discrete r.v., (X,Y) as

$$F(x, y) = \text{Prob}\,[X \leq x, Y \leq y]$$

$$= \sum_{x_i \leq x} \sum_{y_j \leq y} p(x_i, y_j)$$

$$F(\infty, \infty) = P[X \leq \infty, Y \leq \infty] = 1.0$$

Example : Discrete two-d RV

Probability mass function

y \ x	0	1	2	3	4
0	0	0.04	0.05	0.07	0.09
1	0.03	0.04	0.06	0.07	0.08
2	0.02	0.05	0.05	0.07	0.05
3	0.01	0.03	0.05	0.07	0.07

F(3,2) = $\sum_{y=0}^{y=2} \sum_{x=0}^{x=3} p(x, y)$

P[X≤3, Y≤2] = 0+0.04+0.05+0.07+0.03+0.04+0.06+0.07+
0.02+.0.05+.0.05+0.07+0.01+0.03+0.05+0.07
= 0.55

Joint pdf of (X, Y)

- For a continuous r.v. (X, Y), we define the joint probability density function, f(x, y), as
  1) $f(x, y) \geq 0$
  2) $$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y)\,dx\,dy = 1 \longrightarrow$$ This states that the total volume under the surface given by f(x, y) is 1.0
- The joint pdf, f(x, y) is not a probability.
- For small Δx, Δy (+ve), $f(x, y)\,\Delta x\,\Delta y$ is approximately equal to $P[x \leq X \leq x + \Delta x, y \leq Y \leq y + \Delta y]$

Joint cdf of (X, Y)

- The joint cumulative distribution function F(x, y) of the two dimensional random vector (x, y) is defined as

  $F(x, y) = P[X \leq x, Y \leq y]$

$$= \int_{-\infty}^{y}\int_{-\infty}^{x} f(x, y)\,dxdy$$

It follows from the definition, that

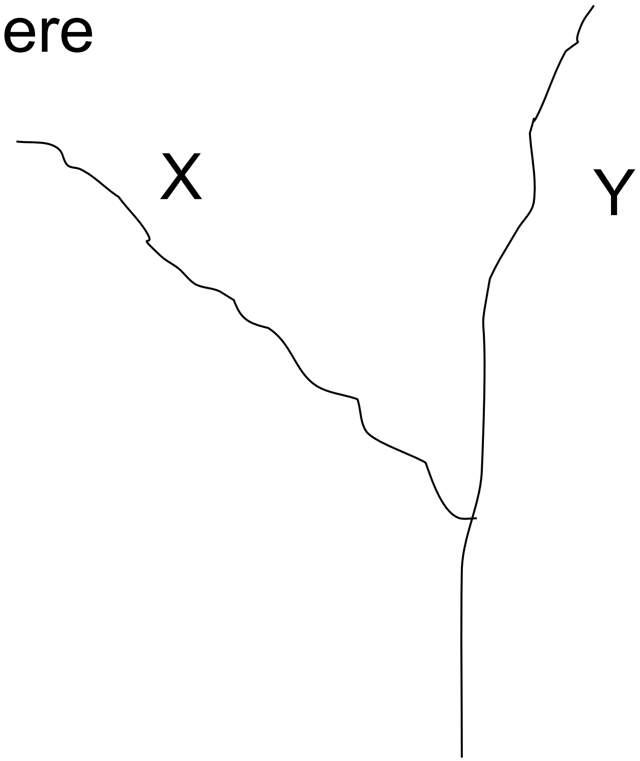
- $F(\infty, \infty)=1.0$
- $F(-\infty, y) = F(x, -\infty) = 0$

Example 1

Flows in two adjacent streams are denoted as a random vector (X, Y) with a joint pdf

$f(x, y) = c \qquad \text{if } 5 \le x \le 10 \; ; \; 4 \le y \le 9$

$\qquad\quad = 0, \qquad \text{elsewhere}$



1. Obtain 'c'
2. Obtain P[X $\geq$ Y]

Example 1 (contd)

1. To determine 'c' $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dx\,dy = 1$

$$\int_{4}^{9}\int_{5}^{10} c\,dx\,dy = 1$$

$$c\int_{4}^{9}\left[x\right]_{5}^{10} dy = 1$$

$$5c\left[y\right]_{4}^{9} = 1$$

$$25c = 1 \Rightarrow c = \frac{1}{25}$$

Example 1 (contd)

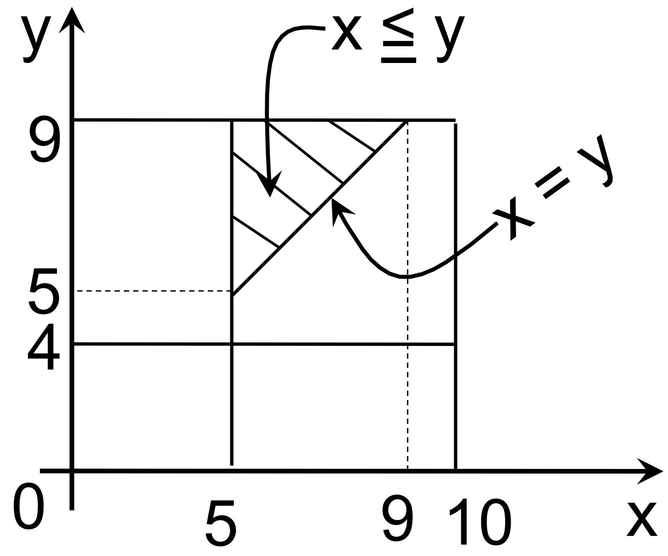
2. P[X $\geq$ Y] = 1-P[X $\leq$ Y]

$$= 1 - \int_5^9 \int_5^y f(x, y)\, dx dy$$

$$= 1 - \frac{1}{25} \int_5^9 \int_5^y dx dy$$

$$= 1 - \left\{ \frac{1}{25} \int_5^9 (y-5) dy \right\}$$

$$= 1 - \left\{ \frac{1}{25} \left[\frac{y^2}{2} - 5y \right]_5^9 \right\}$$



Example 1 (contd)

$$= 1 - \left\{ \frac{1}{25} \left[\frac{9^2}{2} - 5 \times 9 - \frac{5^2}{2} + 5 \times 5 \right] \right\}$$

$$= 1 - 0.32$$

$$= 0.68$$

P[X $\geq$ Y] = 0.68

Example 2

Consider the joint pdf

$$f(x, y) = c(x^2+y^2) \qquad \begin{matrix} 0 \leq x \leq 1 \\ 0 \leq y \leq 1 \end{matrix}$$

$$= 0, \text{ elsewhere}$$

Obtain

1. the constant 'c'
2. $F(x, y)$
3. $P[X \leq 1/2, Y \leq 3/4]$
4. $P[X \geq Y]$
5. $P[X+Y \geq 1]$

Example 2 (contd)

1. To obtain 'c', $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dx\,dy = 1$

$$\int_0^1\int_0^1 c(x^2+y^2)\,dx\,dy = 1$$

$$\int_0^1 c\left[\frac{x^3}{3}+xy^2\right]_0^1 dy = 1$$

$$\int_0^1 c\left[\frac{1}{3}+y^2\right] dy = 1$$

$$c\left[\frac{y}{3}+\frac{y^3}{3}\right]_0^1 = 1 \Rightarrow \frac{2c}{3} = 1 \Rightarrow c = 3/2$$

Example 2 (contd)

2. F(x, y) = $\int_0^x\int_0^y f(x,y)\,dy\,dx = \int_0^x\int_0^y \frac{3}{2}\left(x^2+y^2\right)dy\,dx$

$$= \int_0^x \frac{3}{2}\left(x^2y+\frac{y^3}{3}\right)_0^y dx$$

$$= \int_0^x \frac{3}{2}\left(x^2y+\frac{y^3}{3}\right)dx$$

$$= \frac{3}{2}\left[\frac{x^3y}{3}+\frac{xy^3}{2}\right]_0^x = \frac{x^3y+xy^3}{2}$$

$$F(x,y) = \frac{x^3y+xy^3}{2}$$

$0 \leq x \leq 1$
$0 \leq y \leq 1$

Example 2 (contd)

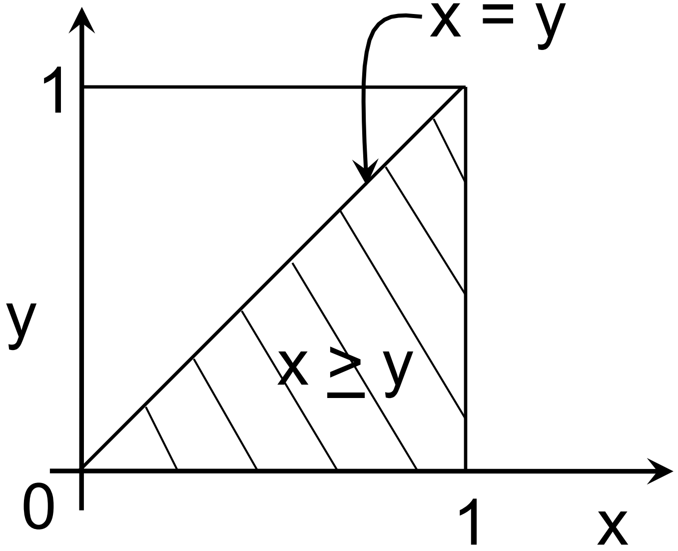
3. P[X ≤ 1/2, Y ≤ 3/4] $= \dfrac{\left(\frac{1}{2}\right)^3 \times \frac{3}{4} + \frac{1}{2} \times \left(\frac{3}{4}\right)^3}{2}$

$$= \frac{39}{256}$$

=0.152

Example 2 (contd)

4. P[Y $\geq$ X]



Limits x → 0 to y

y → 0 to 1

$$P[Y \geq X] = \int_0^1 \int_0^y \frac{3}{2}\left(x^2 + y^2\right) dx dy$$

$$= \int_0^1 \left(\frac{x^3}{2} + \frac{3xy^2}{2} \right)_y^1 dy = \int_0^1 \left(\frac{1}{2} + \frac{3y^2}{2} - 2y^3 \right) dy$$

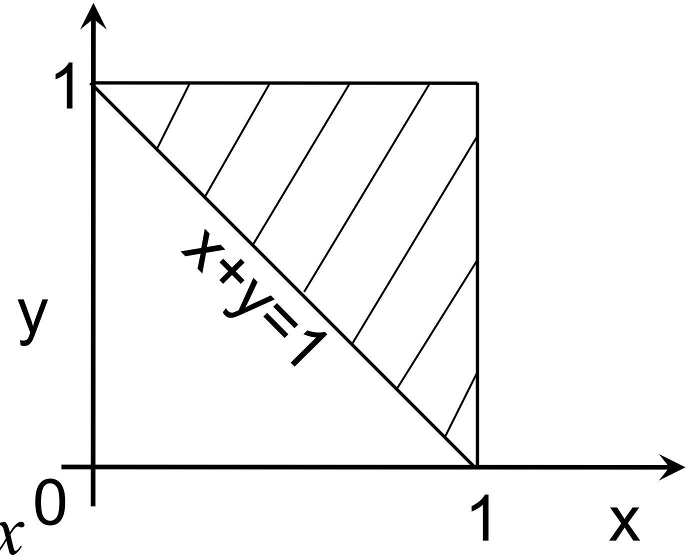
$$= \left[\frac{y}{2} + \frac{y^3}{2} - \frac{y^4}{2} \right]_0^1 = 1/2$$

Example problem-2

5. P[X+Y $\geq$ 1]

Limits x $\rightarrow$ 0 to 1

y $\rightarrow$ 1-x to 1



$$P[X+Y \geq 1] = \int_0^1 \int_{1-x}^1 \frac{3}{2}\left(x^2 + y^2\right) dy dx$$

$$= \int_0^1 \left(\frac{3x^2 y}{2} + \frac{y^3}{2} \right)_{1-x}^{1} dx = \int_0^1 \left(\frac{3x}{2} - \frac{3x^2}{2} + 2x^3 \right) dx$$

$$= \left[\frac{3x^2}{4} - \frac{x^3}{2} + \frac{y^4}{2} \right]_0^1 = 3/4$$

Marginal Probability Distribution

- We have seen f(x, y) as a joint probability distribution
- In discrete case, p(x, y) = P[X = x, Y = y] indicates prob [X=x AND Y=y].
- Consider the following distribution as in the previous numerical example

Marginal Probability Distribution

Marginal distribution of Y

y \ x	0	1	2	3	4	Sum
0	0	0.04	0.05	0.07	0.09	0.25
1	0.03	0.04	0.06	0.07	0.08	0.28
2	0.02	0.05	0.05	0.07	0.05	0.24
3	0.01	0.03	0.05	0.07	0.07	0.23
Sum	0.06	0.16	0.21	0.28	0.29	1.00

Marginal distribution of X

e.g., $P[X \le 3] = 0.06+0.16+0.21+0.28 = 0.71$

Marginal Probability Distribution

- An element in the body of the table indicates $P[X = x_i, Y = y_j]$.
- The marginal totals give $P[Y = y_j]$ and $P[X = x_i]$ resply.
- For example, if we are interested in P[Y = 0], this is given by marginal sum as 0.25.
- Since the event P[Y = 0] can occur with X=0, X=1,…….. X=5. we have P[Y=0, X=0 OR Y=0, X=1 OR …..]

P[Y = 0] = P[Y=0, X=0]+P[Y=0, X=1]+ P[Y=0, X=2]+……
……………… P[Y=0, X=5]

This indicates P[Y=0] irrespective of the value of X

Marginal Probability Distribution

- In general, we may write as

  $p(x_i) = P[X=x_i]$

  $= P[X=x_i, Y=y_1 \text{ or } X=x_i, Y=y_2 \text{or}\ldots\ldots\ldots]$

  $$= \sum_{j=1}^{\infty} p(x_i, y_j)$$

- The function $p(x_i)$ for i=1,2,…. is called the marginal distribution of X.
- Analogously we define $q(y_j) = \sum_{i=1}^{\infty} p(x_i, y_j)$ ∀ j, as the marginal distribution of Y.

Marginal Density Functions

- In the continuous case, we proceed as follows
  - Let f(x, y) denote the joint pdf of (X, Y).
  - We define g(x) and h(y) as the marginal probability density functions of X & Y respectively as

$$g(x) = \int_{-\infty}^{\infty} f(x, y)dy, \; h(y) = \int_{-\infty}^{\infty} f(x, y)dx$$

- These marginal pdfs which are infact derived from the joint pdf f(x, y) correspond to the original pdf's of the one-dimensional r.v.s X and Y.

Marginal Density Functions

This may be seen from

P[c ≤ X ≤ d] = P[c ≤ X ≤ d, -∞ ≤ Y ≤ ∞]

$$= \int_c^d \underbrace{\int_{-\infty}^{\infty} f(x,y)\,dy}\,dx$$

$$= \int_c^d g(x)\,dx$$

From the definitions of pdf's, it is thus seen that g(x) is infact the original pdf of the r.v. X

Marginal Density Functions

$$\text{Thus } g(x) = \int_{-\infty}^{\infty} f(x, y)dy \text{ and } F(x) = \int_{-\infty}^{\infty} g(x)dx$$

Similarly for the r.v. Y

That is, starting with the joint pdf f(x, y), we are able to get the pdfs of X & Y respectively.

For discrete case these results may be written as

$$p\left[X = x_i\right] = \sum_{j=1}^{\infty} p(x_i, y_i) \quad \forall \; i$$

$$p\left[Y = y_j\right] = \sum_{i=1}^{\infty} p(x_i, y_i) \quad \forall \; j$$

Example 3

Consider the joint pdf in the previous example

$$f(x, y) = 1/25 \qquad \begin{array}{l} 5 \le x \le 10 \\ 4 \le y \le 9 \end{array}$$

$$= 0, \text{ elsewhere}$$

1. Obtain the marginal density g(x), h(y)
2. Obtain CDF G(x), H(y)
3. $P[X \ge 7]$
4. $P[5 \le Y \le 8]$

Example 3 (contd)

1. To obtain g(x),

$$g(x) = \int_{-\infty}^{\infty} f(x,y)dy \qquad 5 \le x \le 10$$

$$= \int_{4}^{9} \frac{1}{25} dy$$

$$= \left[\frac{y}{25}\right]_{4}^{9} = 1/5$$

$$g(x) = \frac{1}{5} \qquad 5 \le x \le 10$$

Example 3 (contd)

1. To obtain h(y),

$$h(y) = \int_{-\infty}^{\infty} f(x,y)dx \qquad 4 \le y \le 9$$

$$= \int_{5}^{10} \frac{1}{25} dx$$

$$= \left[\frac{x}{25} \right]_5^{10} = 1/5$$

$$h(y) = \frac{1}{5} \qquad 4 \le y \le 9$$

Example 3 (contd)

2. To obtain G(x),

$$G(x) = \int_{-\infty}^{x} g(x)dx \qquad 5 \le x \le 10$$

$$= \int_{5}^{x} \frac{1}{5} dx$$

$$= \left[\frac{x}{5} \right]_{5}^{x}$$

$$G(x) = \frac{x-5}{5} \qquad 5 \le x \le 10$$

Example 3 (contd)

2. To obtain H(y),

$$H(y) = \int_{-\infty}^{y} h(y)dy \qquad 4 \le y \le 9$$

$$= \int_{4}^{y} \frac{1}{5} dy$$

$$= \left[\frac{y}{5} \right]_{4}^{y}$$

$$H(y) = \frac{y-4}{5} \qquad 4 \le y \le 9$$

Example 3 (contd)

3. $P[X \geq 7] = 1 - P[X \leq 7]$

$$= 1 - G(7)$$

$$= 1 - \frac{7-5}{5} = \tfrac{3}{5}$$

4. $P[5 \leq Y \leq 8] = H(8) - H(5)$

$$= \frac{4}{5} - \frac{1}{5} = \tfrac{3}{5}$$

Example 4

Consider the joint pdf

$$f(x, y) = e^{-y} \qquad \begin{matrix} x > 0 \\ y \geq x \end{matrix}$$

1. Obtain the marginal density g(x)
2. $P[X \geq 2]$

Example 4 (contd)

1. To obtain g(x),

$$g(x) = \int_{-\infty}^{\infty} f(x,y)dy \qquad x > 0$$

$$= \int_{x}^{\infty} e^{-y}\, dy$$

$$= \left[-e^{-y}\right]_{x}^{\infty} = e^{-x}$$

$$G(x) = \int_{0}^{x} e^{-x} dx = \left[-e^{-x}\right]_{0}^{x}$$

$$= 1 - e^{-x} \qquad \text{x > 0}$$

Example 4 (contd)

2. $P[X \geq 2] = 1 - P[X \leq 2]$

$$= 1 - \left(1 - e^{-2}\right)$$

$$= e^{-2}$$

Conditional Distribution

- A marginal distribution is the distribution of one variable regardless of the value of the second variable
- A joint distribution is the simultaneous occurrence of the given values of the two variables
- The distribution of one variable with conditions placed on the second variable is called conditional distribution. For example,
  - Distribution of X, given that $Y=y_0$ or distribution of Y given that $c \leq X \leq d$ etc.

Conditional Distribution

Definition: (X, Y) is a continuous two dimensional r.v. with a joint pdf of f(x, y).

- Let g(x) and h(y) be the marginal pdfs of X and Y respectively
- The conditional pdf of X given Y = y is defined as

$$g\left(x/y\right) = \frac{f(x, y)}{h(y)} \quad h(y) > 0$$

↑ Read as x given y

Conditional Distribution

- The conditional pdf of Y given X = x is defined as

$$h\left(y/x\right) = \frac{f(x,y)}{g(x)} \quad g(x) > 0$$

- The conditional pdf of X given Y∈R and conditional pdf of Y given X∈R is defined as

$$g\left(x/y\in R\right) = \frac{\int_R f\left(x,y\right)dy}{\int_R h\left(y\right)dy}, \quad h\left(y/x\in R\right) = \frac{\int_R f\left(x,y\right)dx}{\int_R g\left(x\right)dx}$$

Conditional Distribution

The conditional pdfs g(x/y) and h(y/x) satisfy all conditions for a pdf.

For a given y, g(x/y)>0, as both f(x,y) and h(y) are positive.

$$\int_{-\infty}^{\infty} g\left(x/y\right) = \int_{-\infty}^{\infty} \frac{f(x,y)}{h(y)}\ dx$$

$$= \frac{1}{h(y)} \int_{-\infty}^{\infty} f(x,y) dx$$

$$= \frac{h(y)}{h(y)} = 1.0$$

Cumulative conditional distributions

$$G\left(x/y\right) = \int_{-\infty}^{x} g\left(x/y\right)\ dx, \quad H\left(y/x\right) = \int_{-\infty}^{y} h\left(y/x\right)\ dy$$

Example 5

Consider the joint pdf

$$f(x, y) = \frac{x\left(1+3y^2\right)}{4} \qquad 0 \leq x \leq 2,\; 0 \leq y \leq 1$$

$= 0$, elsewhere

1. Obtain h(y/x)
2. $P[1/2 \leq Y \leq 1 / X=1]$
3. $P[Y \leq 3/4 \;/\; X \leq 1]$

Example 5 (contd)

1. To obtain h(y/x), g(x) is first obtained

$$g(x) = \int_{-\infty}^{\infty} f(x,y)dy \qquad 0 \le x \le 2$$

$$= \int_0^1 \frac{x\left(1+3y^2\right)}{4} dy$$

$$= \frac{1}{4}\left[xy + y^3\right]_0^1$$

$$g(x) = \frac{\left(x+1\right)}{4} \qquad 0 \le x \le 2$$

Example 5 (contd)

$$h\left(y/x\right) = \frac{f\left(x, y\right)}{g\left(x\right)}$$

$$= \frac{x\left(1+3y^2\right)}{x+1}$$

2. P[1/2 ≤ Y ≤ 1/X=x] $= \int_{1/2}^{1} \frac{x}{x+1}\left(1+3y^2\right)dy$

$$= \frac{x}{x+1}\left[y + y^3\right]_{1/2}^{1}$$

$$= \frac{11}{8}\left(\frac{x}{x+1}\right)$$

Example 5 (contd)

2. $P[1/2 \le Y \le 1/X=1] = \frac{11}{8}\left(\frac{1}{1+1}\right) = {}^{11}\!/_{16}$

3. To obtain $P[Y \le 3/4 / X \le 1]$,

$$h\left(y/x \in R\right) = \frac{\int_R f\left(x,y\right)dx}{\int_R g\left(x\right)dx}$$

$$\Rightarrow h\left(y/x \le 1\right) = \frac{\int_0^1 f\left(x,y\right)dx}{\int_0^1 g\left(x\right)dx}$$

Example 5 (contd)

$$\int_0^1 f(x,y)dx = \int_0^1 \frac{x\left(1+3y^2\right)}{4}dx$$

$$= \frac{1+3y^2}{4}\left[\frac{x^2}{2}\right]_0^1$$

$$= \frac{1+3y^2}{8}$$

$$\int_0^1 g(x)dx = \int_0^1 \frac{\left(x+1\right)}{4}dx$$

$$= \frac{1}{4}\left[\frac{x^2}{2}+x\right]_0^1 = 3/8$$

Example 5 (contd)

$$h\left(y/x \le 1\right) = \frac{1+3y^2}{3}$$

$$h\left(y \le 3/4/x \le 1\right) = \int_0^{3/4} \frac{1+3y^2}{3} dy$$

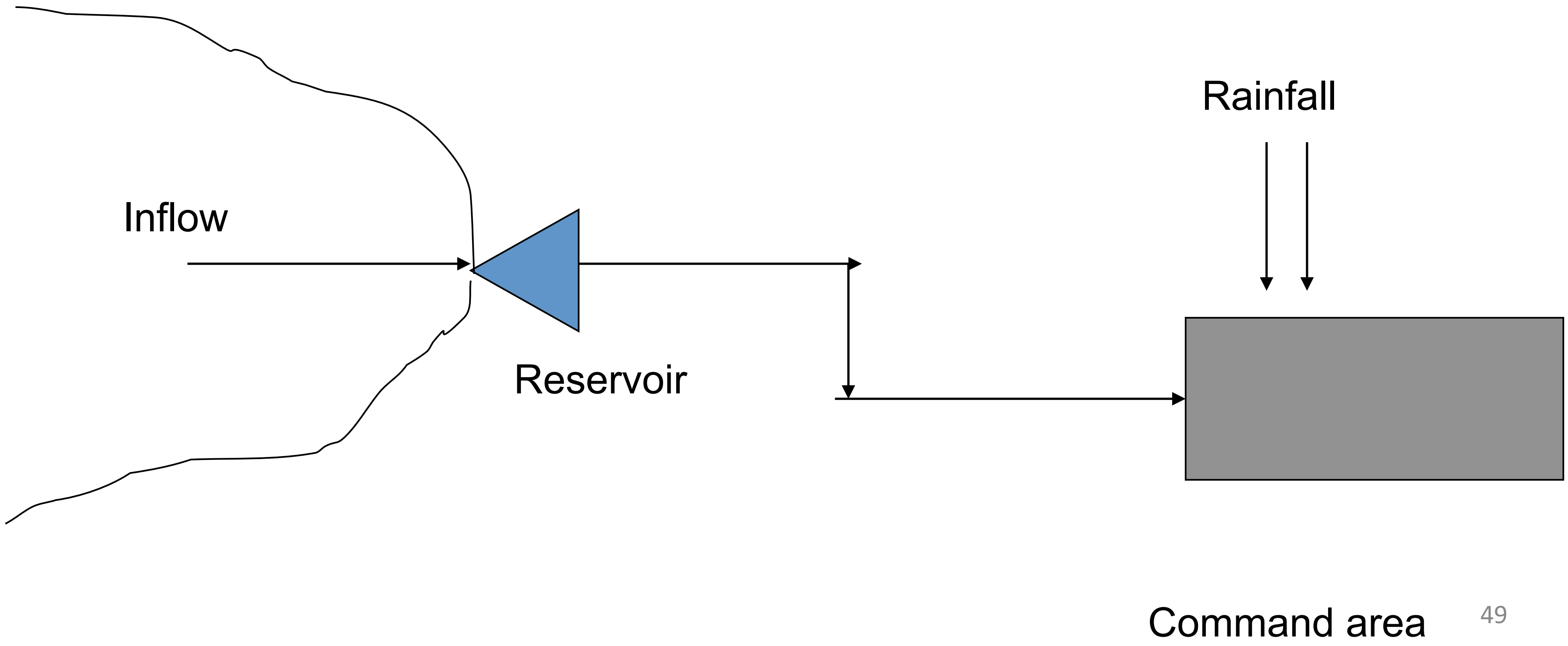
$$= \frac{1}{3}\left[y + y^3\right]_0^{3/4}$$

$$= \frac{1}{3}\left[\frac{3}{4} + \frac{27}{64}\right] = 75/192$$

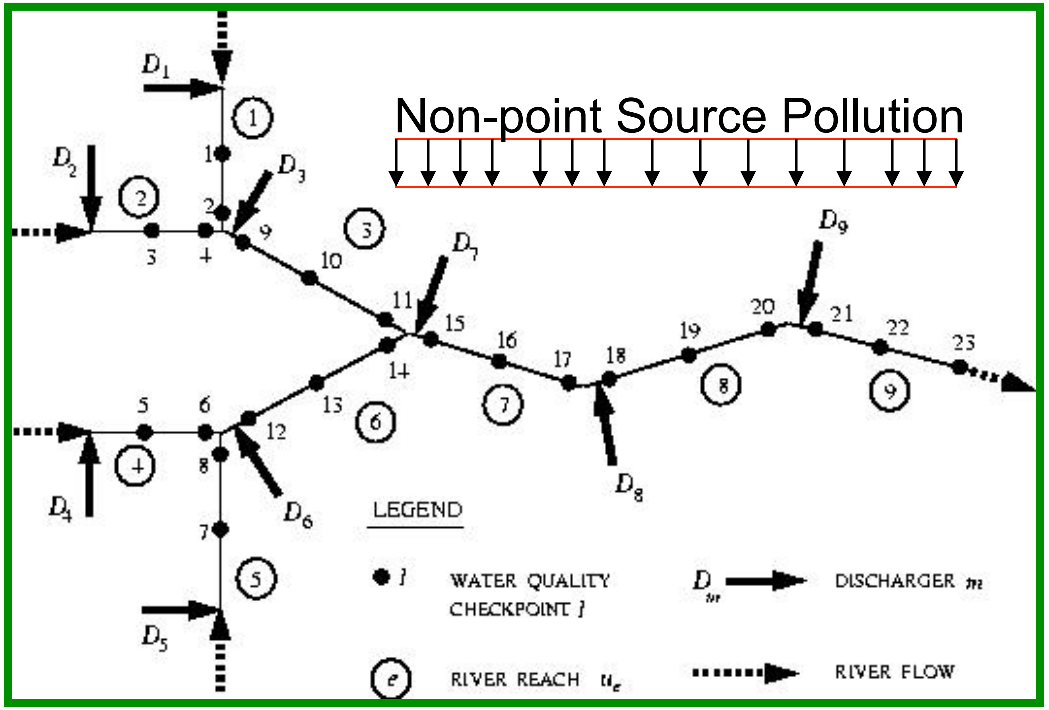
Independence of two random variables

- Intuitively, X and Y are independent r.v.s if the distribution of one r.v. does not in any way influence distribution of the other r.v.
- Independence is a very useful assumption for hydrologic analysis in many situations. However, physically the assumption must have a sound basis.

- As an example, inflow to a reservoir (X) and the rainfall in the command area (Y) may be taken as independent, if the command area is far removed from the reservoir.



- In water quality problems, for example, pollutant load (X) and stream flow (Y) may be treated as independent variables.



Independent R.V.

- When two r.v.s are independent, g(x/y)=g(x)
  - Distribution of x given y is independent of y and hence the original pdf itself gives the conditional pdf

$$g\left(x/y\right) = \frac{f(x,y)}{h(y)}$$

$$g(x) \quad = \frac{f(x,y)}{h(y)}$$

$$f(x,y) = g(x).h(y)$$

Independent R.V.

- The random variables X and Y are stochastically independent if and only if their joint density is equal to the product of their marginal densities.
- For discrete case, the two r.v.s are independent if and only if

$$p(x_i, y_j) = p(x_i) \,.\, p(y_j) \;\forall\; i,j$$