UNIVERSITAS GADJAH MADA
DEPARTMENT OF CIVIL AND ENVIRONMENTAL ENGINEERING
BACHELOR IN CIVIL ENGINEERING

Statistics and Probability

Random Variables

Random variables

- Definition
  - A random variable is a variable whose value is unknown or a function that assigns values to each of an experiment’s outcomes
  - A random variable is a function defined in a sample space
- Type
  - Discrete random variables
  - Continuous random variables
- Example
  - Number of rainy days in a year → discrete random variable
  - Precipitation in a year → continuous random variable

Random variables

- Notation
  - X → variable random
  - x → the value of a random variable
- Function
  - A function of random variables is a random variable
  - If X is a variable random, then $Z = f(X)$ is a random variable

Discrete random variables

$X =$ random variables

$= x_1, x_2, x_3, \ldots, x_n$

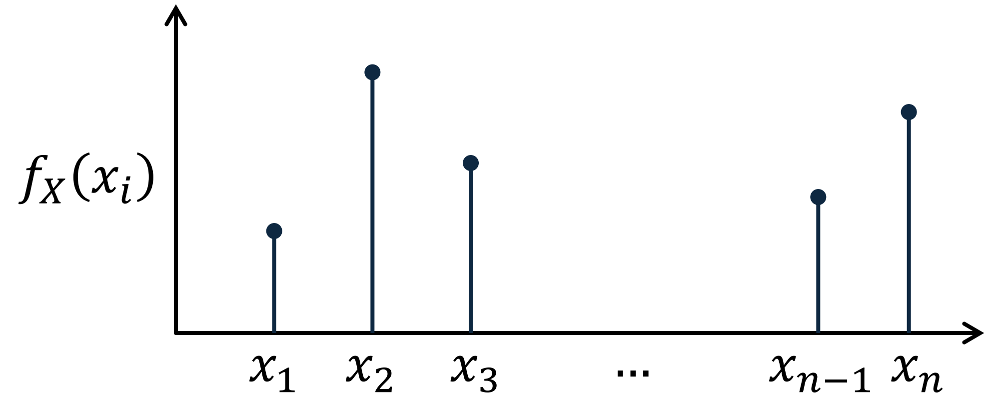
$f(x_1)$

$f(x_2)$

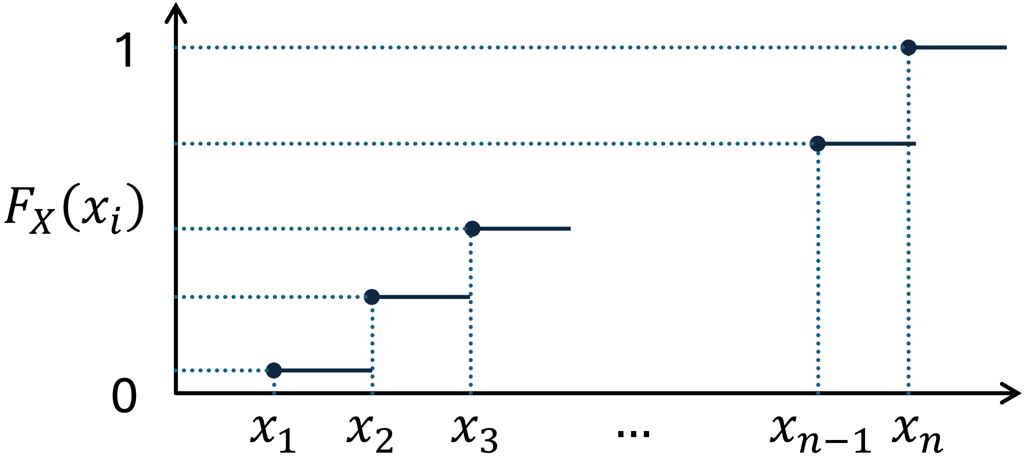
$f(x_3)$

$f(x_n)$

$$\text{probability} = \sum_{i=1}^{n} f_X(x_i) = 1$$



discrete probability distribution



discrete cumulative probability distribution

⇩

probability

$x \leq x_i$

- Cumulative probability distribution of a random variable X for $X = x$

$$F_X(x) = \sum_{x_i \le x} f(x_i)$$

- Probability distribution of a random variable X for $X = x$

$$f_X(x_i) = F_X(x_i) - F_X(x_{i-1})$$

- Relative frequency

$$f_{x_i} = F_{x_i} - F_{x_{i-1}}$$

- Probability

$$f_X(x_i) = F_X(x_i) - F_X(x_{i-1})$$

- Cumulative relative frequency

$$F_{x_i} = \sum_{j=1}^{i} f_{x_j}$$

$$F_X(x) = \sum_{x_i \le x} f_X(x_i)$$

Continuous random variables

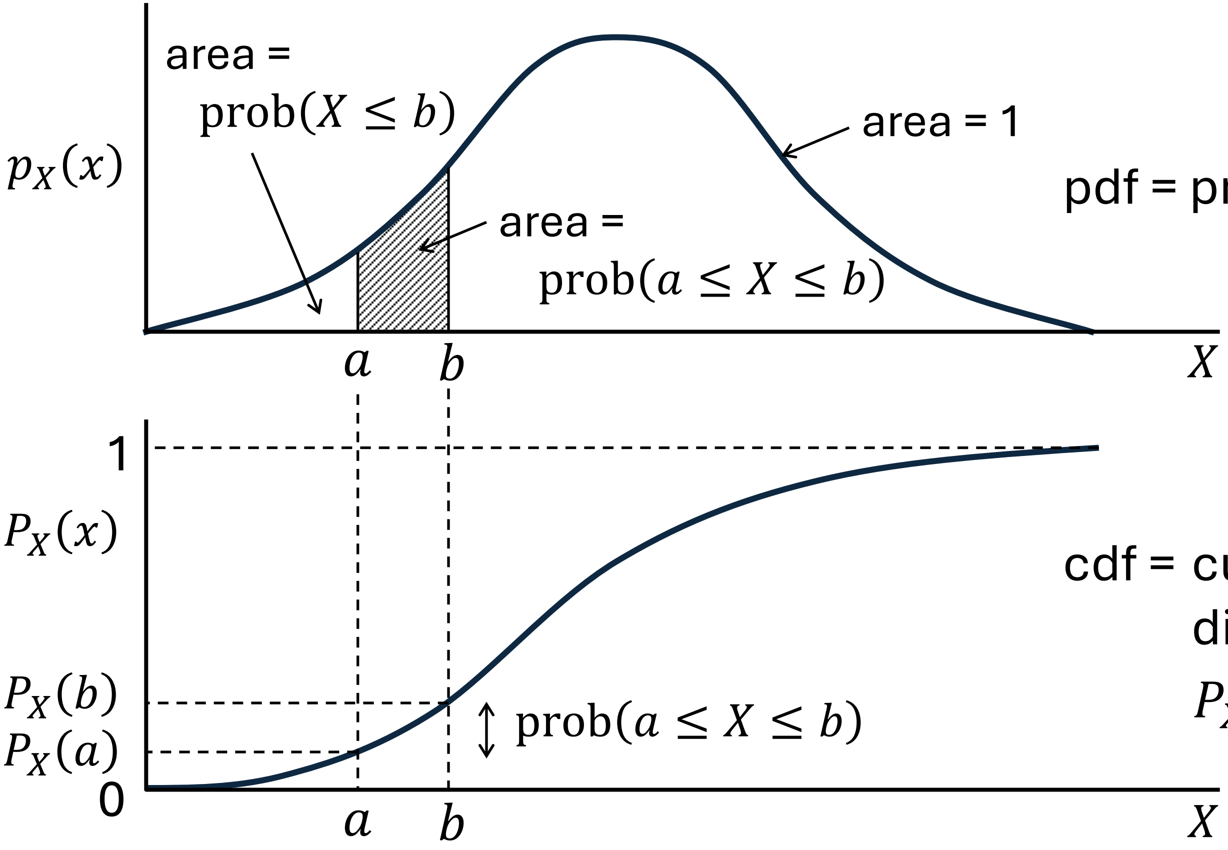
- Probability

$$\text{prob}(A) = \frac{n_i}{n} = f_{x_i}$$

n_i is the number of data in the i-th class
n is the number of data

- Thus, f_{x_i} can be read as the estimate of the probability
  - $f_{x_i} \longrightarrow$ estimate of $\text{prob}(A)$
  - frequency histogram $\longrightarrow$ estimate of probability distribution
  - cumulative frequency $\longrightarrow$ estimate of cumulative probability distribution

continuous random variables are treated as if they are discrete ones



pdf = probability density function

cdf = cumulative probability distribution function

$P_X(x) = \text{prob}(X \le x)$

$p_X(x)$ = probability density function of a continuous random variable

$P_X(x)$ = cumulative probability distribution function

$$\Rightarrow P_X(x) = \text{prob}(X \le x) \Rightarrow dP_X(x) = p_X(x)dx \Rightarrow P_X(x) = \int_{-\infty}^{x} p_X(t)dt$$

Some probability properties

(1) $p_X(x) \geq 0, \forall x$

(2) $\int_{-\infty}^{+\infty} p_X(x)dx = 1$

(3) $P_X(-\infty) = 0$

(4) $P_X(+\infty) = 1$

(5) $\text{prob}(a \leq X \leq b) = \int_a^b p_X(t)dt = P_X(b) - P_X(a)$

(6) $\text{prob}(X = c) = \int_c^c p_X(t)dt = P_X(c) - P_X(c) = 0$

$\Downarrow$

$\text{prob}(a \leq X \leq b) = \text{prob}(a < X \leq b) = \text{prob}(a \leq X < b)\text{prob}(a < X < b)$

Example #1

- Below is the pdf of a random variable X

$$p_X(x) = \begin{cases} x/2 & \text{if } 0 < x < 2 \\ 0 & \text{else} \end{cases}$$

- Draw the pdf
- Show that $\text{prob}(0 < X < 2) = 1$
- Find $\text{prob}(X < 1.5) = P_X(1.5)$
- Find $\text{prob}(0.5 < X < 1.5)$

Example_#1_Random_Variables.pdf

Example #2

- Annual series of maximum daily rainfall at an ARR (Automatic Rainfall Recorder), H mm, shows that its pdf can be expressed as follow

$$p_H(h) = \begin{cases} \dfrac{1}{75} & \text{if } 0 < h < 50 \\ \dfrac{1}{3750}(100 - h) & \text{if } 50 < h < 100 \\ 0 & \text{else} \end{cases}$$

  - Draw the pdf
  - Find and draw the cdf
  - Find $\text{prob}(40 \text{ mm} < H < 60 \text{ mm})$

Example_#2_Random_Variables.pdf

Random Variables

Bivariate Random Variables

Bivariate distributions

- The previous section discusses univariate distributions
- There exists bivariate distributions
  - We calculate the joint probabilities when we study the behavior of bivariate distributions
  - Joint probabilities $\longrightarrow$ joint pdf

Bivariate distributions

- Joint distributions

$$p_{X,Y}(x,y) = \frac{\partial}{\partial x \partial y} P_{X,Y}(x,y)$$ pdf

$$P_{X,Y}(x,y) = \mathrm{prob}(X < x \wedge Y < y)$$ cdf

$$= \iint_{-\infty}^{+\infty} p_{X,Y}(s,t)dtds$$

Bivariate distributions

- Some properties of bivariate distribution

$P_{X,Y}(x,\infty)$ → cdf of random variable X (univariate)

$P_{X,Y}(\infty,y)$ → cdf of random variable Y (univariate)

$$P_{X,Y}(x,y) > 0$$

$$P_{X,Y}(+\infty,+\infty) = 1$$

$$P_{X,Y}(-\infty,y) = P_{X,Y}(x,-\infty) = 0$$

Marginal distributions

- Two random variables X and Y
  - The behavior of X without considering Y
  - Marginal density (pdf) and cumulative marginal distribution (cdf)

pdf

$$p_{X,Y}(x,y) \longrightarrow p_X(x)$$

$$p_X(x) = \int_{-\infty}^{+\infty} p_{X,Y}(x,t)dt$$

cdf

$$P_{X,Y}(x,y) \longrightarrow P_X(x)$$

$$\begin{aligned} P_X(x) &= P_X(x,\infty) \\ &= \text{prob}(X < x \wedge Y < +\infty) = \text{prob}(X < x) \\ &= \int_{-\infty}^{x}\int_{-\infty}^{+\infty} p_{X,Y}(s,t)dtds = \int_{-\infty}^{x} p_X(s)ds \end{aligned}$$

Marginal distributions

- Two random variables X and Y
  - The behavior of Y without considering X

pdf

$$p_{X,Y}(x,y) \longrightarrow p_Y(y)$$

$$p_Y(y) = \int_{-\infty}^{+\infty} p_{X,Y}(s,y)ds$$

cdf

$$P_{X,Y}(x,y) \longrightarrow P_Y(y)$$

$$\begin{aligned} P_Y(y) &= P_Y(\infty, y) \\ &= \text{prob}(X < +\infty \wedge Y < y) = \text{prob}(Y < y) \\ &= \int_{-\infty}^{+\infty} \int_{-\infty}^{y} p_{X,Y}(s,t)dtds = \int_{-\infty}^{y} p_Y(t)dt \end{aligned}$$

Conditional Distributions

- Two random variables X and Y
  - The behavior of X, which depends on Y
    - The distribution of X for $Y = y_0$
    - The distribution of Y for $x_1 < X < x_2$

$$p_X(x_i|y \in S) = \frac{\int_S \; p_{X,Y}(x,t)dt}{\int_S \; p_Y(t)dt}$$

Conditional probability

$$p_X(x_i|y \in S) = \frac{\int_S \; p_{X,Y}(x,t)dt}{\int_S \; p_Y(t)dt}$$

$$\text{prob}(x \in R|y \in S) = \int_R \; p_{X|Y}(x|y \in S)dx$$

$$p_{X|Y}(x|y = y_0) = \frac{p_{X|Y}(x, y_0)}{p_Y(y_0)}$$

more readable

$$p_{X|Y}(x|y) = \frac{p_{X|Y}(x, y)}{p_Y(y)}$$

Independence

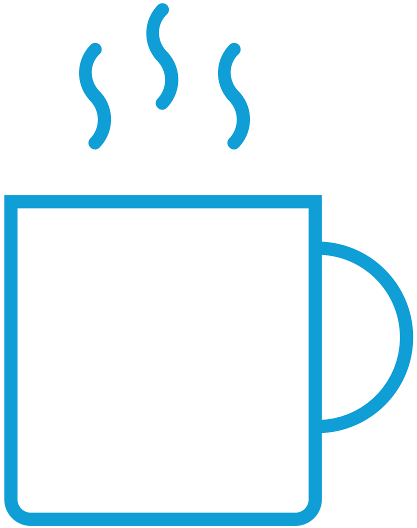
- Two random variables X and Y
  - X and Y are independent if
    - $p_{X|Y}(x|y)$ is not a function of y
    - $p_{X|Y}(x|y) = p_X(x)$
  - Joint probabilities
    - Multiplication of two marginal densities
    - $p_{X,Y}(x,y) = p_X(x) \cdot p_Y(y)$

Example

		Air temperature, T°C					
		22—24	24—26	26—28	28—30	30—32	32—34
Air humidity, H%	0—20	2	4	6	2	2	1
	20—40	4	8	12	30	6	9
	40—60	5	15	30	60	30	20
	60—80	3	7	9	25	17	11
	80—100	1	0	2	12	8	3

Example

- From the data, find
  - joint pdf
  - marginal pdf and cdf of the air temperature
  - marginal pdf and cdf of the air humidity
  - probability of air temperature be in the range of 28°C to 30°C
  - probability of air temperature be in the range of 28°C to 30°C when the air humidity is in the range of 60% to 80%



Statistics and Probability

Random Variables