

Chapter 3 – Random Variables and Probability Distributions (I)

Engr 0021

Review

- Chapter 1
 - Introduction
 - Measure of Location
 - Measure of Variability
- Chapter 2
 - Define sample space and measurable events
 - Assign probability measure

Learning Objectives

- Concept of a Random Variable
- Discrete Probability Distribution
- Continuous Probability Distribution
- Joint Probability Distribution
- Marginal Distribution
- Conditional Distribution

Concept of a Random Variable

- Statistical experiment: three electronic components are tested

$$S = \{NNN, NND, NDN, DNN, NDD, DND, DDN, DDD\}$$

- N : Non-defective
- D : Defective
- X – the number of defective items when 3 electronic components are tested

Random Variable

- For a given sample space S of some experiment, a random variable is any rule that associates a number with each outcome in S .
- Notation:
 - X : random variable
 - x : one realization or one observed value for random variable X
- Example: die toss, number of defects in a sample of 100 productions, number of experimental trials before success, depth measurement of a lake, pH of a chemical compound, number of people who will vote for a particular candidate, weight of a coil of steel, length of a steel beam, diameter of a washer, etc.

Random Variable: Examples

- Consider an experiment that consists of tossing two coins.
- A die is thrown until 5 occurs.
- Components are arriving from a production line and they are stipulated to be defective or not defective.
- Record the successive speeders spotted by a radar unit.

Random Variable: Examples

- Consider an experiment that consists of tossing two coins.
 - Let X equal the number of heads observed.
- A die is thrown until 5 occurs.
 - Let X be the number of times the die is thrown before 5.
- Components are arriving from a production line and they are stipulated to be defective or not defective.
 - Let X be the number of defective parts.
- Record the successive speeders spotted by a radar unit.
 - Let X be a random variable defined by waiting time, in hours between successive speeders spotted by a radar unit.

Discrete and Continuous Sample Space

- If a sample space contains a finite number of possibilities or an unending (infinite) sequence with as many elements as there are whole integers, it is called **discrete sample space**.
- If a sample space contain an infinite number of possibilities equal to the number of points on a line segment, it is called a **continuous sample space**.

Types of Random Variables

- A **discrete** random variable is a random variable whose possible values either constitute a finite set or else can be listed in an infinite sequence in which there is a first element, a second element, and so on (“countable” infinite).
- A random variable is **continuous** if both of the following apply:
 - Its set of possible values consists either of all numbers in a single interval on the number line (possibly infinite in extent, e.g., from $-\infty$ to ∞) or all numbers in a disjoint union of such intervals (e.g., $[0, 10] \cup [20, 30]$).
 - No possible value of the variable has positive probability, i.e., $P(X = c) = 0$ for all possible value c .

Examples: Discrete or Continuous?

Are the following random variables discrete or continuous?

- The total number of points scored in a basketball game
- The shelf life of a particular drug
- The length of a two-year old black bass
- The number of aircraft near-collisions last year

Probability Distribution

- The Probability Distribution for a random variable X tells us how the total probability of 1 for the sample space S is distributed among each of the mutually exclusive simple events (or outcomes for X) that describe the sample space.

Discrete Probability Distribution Example

- Consider an experiment that consists of tossing two coins and let X equal the number of heads observed. Find the probability distribution for X .

Simple Event	Coin 1	Coin 2	Outcome	$p(E_i)$	x
E_1	H	H	HH		
E_2	H	T	HT		
E_3	T	H	TH		
E_4	T	T	TT		

X	$p(x)$
0	
1	
2	

Discrete Probability Distribution

The set of ordered pairs $(x, f(x))$ is a probability function, probability mass function, or probability distribution of the discrete random variable X if, for each possible outcome x , denoted as pmf:

1. $f(x) \geq 0$
2. $\sum_x f(x) = 1$
3. $P(X = x) = f(x)$

Example

- A probability distribution for a random variable X :

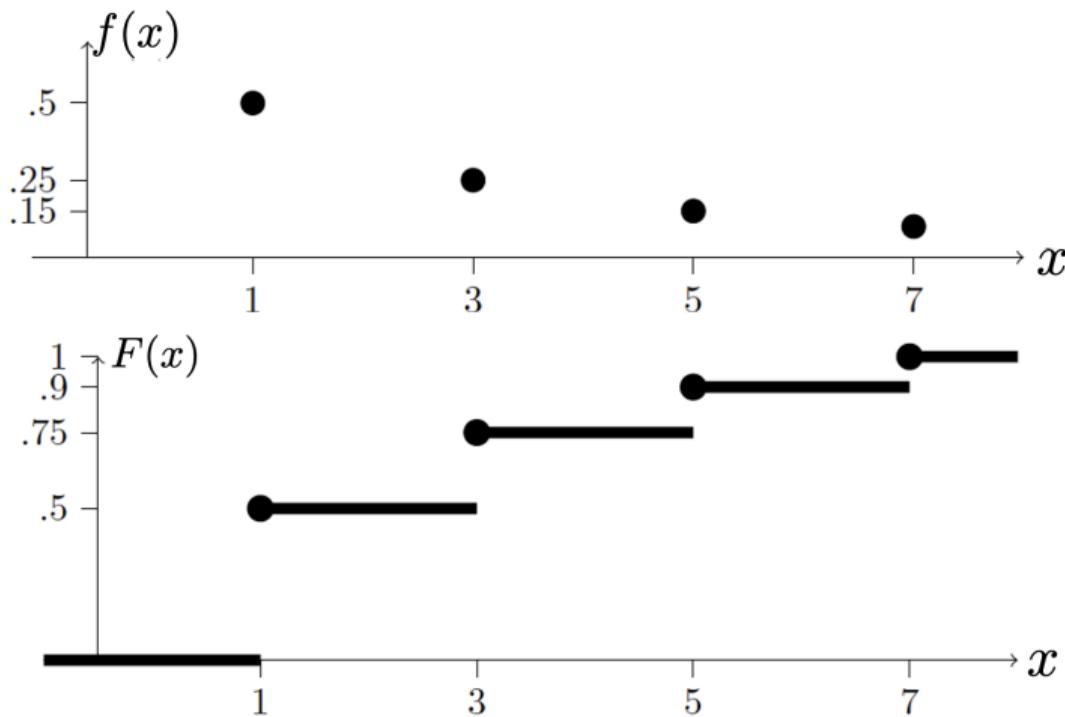
x	-8	-3	-1	0	1	4	6
$P(X = x)$	0.13	0.15	0.17	0.20	0.15	0.11	0.09

- Find
 - $P(X \leq 0)$
 - $P(-3 \leq X \leq 1)$

Marbles Example

A box contains 8 red marbles and 3 green marbles. Six marbles are drawn at random without replacement. (How many outcomes result in all 3 green marbles being drawn?)
Find the probability distribution for the number of green marbles drawn.

Probability mass function plot



The Cumulative Distribution Function (Discrete)

The cumulative distribution function $F(x)$ of discrete random variable X with probability distribution $f(x)$ is

$$F(x) = P(X \leq x) = \sum_{t \leq x} f(t), \quad -\infty < x < \infty.$$

Example

- A probability distribution for a random variable X :

x	-8	-3	-1	0	1	4	6
$P(X = x)$	0.13	0.15	0.17	0.20	0.15	0.11	0.09

- Write the CDF of X

Proposition

- For any two numbers a and b with $a \leq b$,

$$P(a \leq X \leq b) = F(b) - F(a^-)$$

- ' a^- ' represents the largest possible X value that is strictly less than a
- Note: for integers (as possible values for x):

$$P(a \leq X \leq b) = F(b) - F(a - 1)$$

Example

- Suppose a random variable has the following distribution

x	0	1	2	3	4
$f(x)$	0.4	0.1	0.1	0.1	0.3

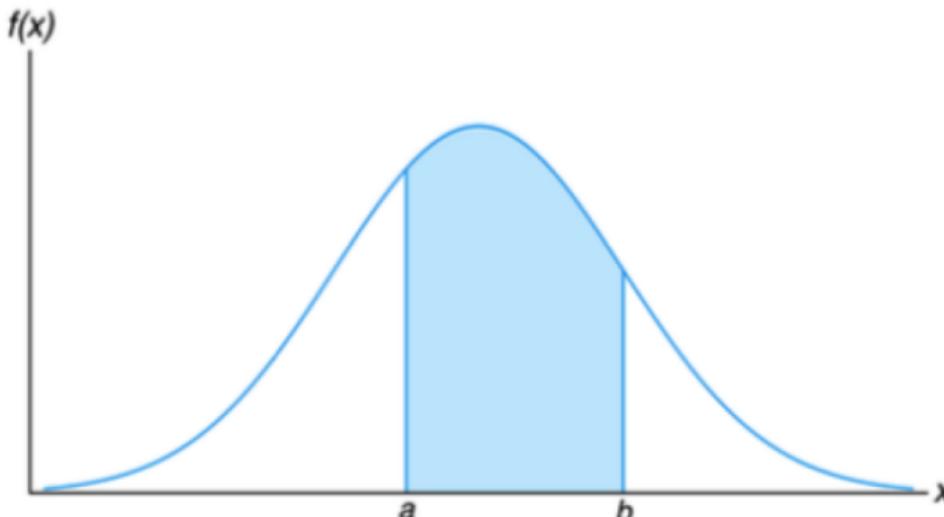
- Find the CDF
- Find $P(X \leq 2)$
- Find $P(X < 2)$
- Find $P(2 \leq X \leq 4)$
- Find $P(X \neq 0)$

Continuous Random Variable

- Continuous random variable can assume infinitely many value corresponding to points in a particular interval.
- A random variable X is continuous if its set of possible values is an entire interval of number (If $A < B$, then any number between A and B is possible.)
- If the measurement scale of X can be subdivided to any extent desired, then a variable is continuous; if it cannot, the variable is discrete.

Continuous Probability Distribution

- X – random variable whose values are heights of all people over 21 years of age.



$$P(a < X < b)$$

Continuous Probability Distribution

The function $f(x)$ is a probability density function, denoted as pdf, for the continuous random variable X , defined over the set of real numbers, if

1. $f(x) \geq 0$, for all $x \in \mathbb{R}$
2. $\int_{-\infty}^{\infty} f(x) dx = 1$
3. $P(a < X < b) = \int_a^b f(x) dx$

The Cumulative Distribution Function (Continuous)

The cumulative distribution function $F(x)$ of a continuous random variable X with density function $f(x)$ is

$$F(x) = P(X \leq x) = \int_{-\infty}^x f(t) dt, \quad -\infty < x < \infty$$

Example

- Suppose the pdf of the magnitude X of a dynamic load on a bridge (in newtons) is

$$f(x) = \begin{cases} \frac{1}{8}x + \frac{3}{8}, & 0 \leq x \leq 2 \\ 0, & \text{otherwise} \end{cases}$$

- Find $F(x) = \int_{-\infty}^x f(y) dy$

Example

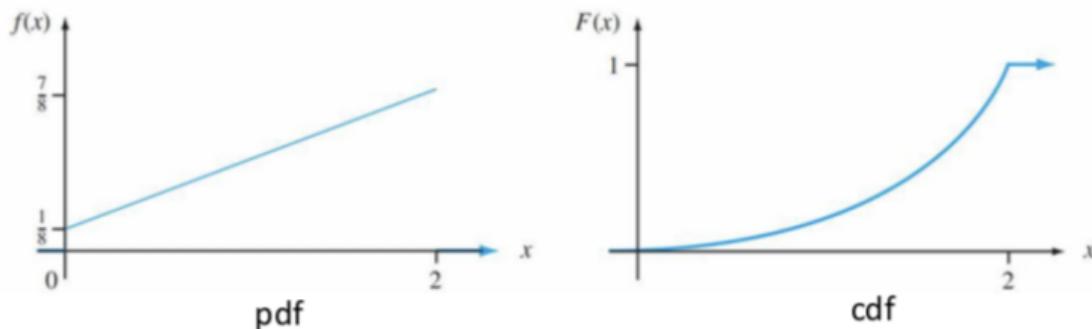


Figure: pdf and CDF

- Find the probability that the load is between 1 and 1.5
- Find the probability that the load exceeds 1

Example

- Consider the density function

$$f(x) = \begin{cases} \frac{k}{\sqrt{x}}, & 0 < x < 1 \\ 0, & \text{otherwise} \end{cases}$$

- Evaluate k
- Find $F(x)$ and use it to evaluate $P(0.3 < X < 0.6)$

Obtaining $f(x)$ from $F(x)$

- For X discrete, the pmf is obtained from the CDF by taking the difference between two $F(x)$ values. The continuous analog of a difference is a derivative.
- The following result is a consequence of the Fundamental Theorem of Calculus.

Proposition: If X is a continuous random variable with pdf $f(x)$ and CDF $F(x)$, then at every x at which the derivative $F'(x)$ exists,

$$F'(x) = f(x).$$

Review

- Concept of a random variable
- Focus on univariate random variable so far
- Discrete probability distributions
 - pmf
 - CDF
- Continuous distributions
 - pdf
 - CDF

This Lecture

- Joint Probability Distribution (Multiple random variables)
- Marginal Distribution
- Conditional Distribution
- Statistical Independence

Joint Probability Distributions

- Interested in the probability multiple random variables take on particular values
- Examples:
 - demand for three different products next period
 - number of different birthdays for the women and men in the room
 - height and weight of an animal one year from now
- There are many experimental situations in which more than one random variable is of interest
- Joint distributions can be discrete or continuous (or neither); we focus on discrete and continuous cases

Example: Joint pmf Table (Insurance Deductibles)

- X = homeowner's deductible, Y = auto deductible.
- Joint probability table $f(x, y)$:

$f(x, y)$	$y = 0$	$y = 100$	$y = 200$
$x = 100$	0.20	0.10	0.20
$x = 250$	0.05	0.15	0.30

Joint Probability Distribution (Discrete)

- The function $f(x, y)$ is a **joint probability distribution** (joint pmf) of discrete random variables X and Y if:
 1. $f(x, y) \geq 0$ for all x, y
 2. $\sum_x \sum_y f(x, y) = 1$
 3. $P(X = x, Y = y) = f(x, y)$
- For any region A in the xy -plane:

$$P((X, Y) \in A) = \sum \sum_{(x,y) \in A} f(x, y).$$

Example 3.14

- Two ballpoint pens are selected at random from a box with:
3 blue, 2 red, 3 green (8 total).
- X = number of blue pens selected, Y = number of red pens selected.
- Find the joint pmf $f(x, y)$.

Example 3.14 (Joint pmf)

$f(x, y)$	$y = 0$	$y = 1$	$y = 2$
$x = 0$	$\frac{3}{28}$	$\frac{3}{14}$	$\frac{1}{28}$
$x = 1$	$\frac{9}{28}$	$\frac{3}{14}$	0
$x = 2$	$\frac{3}{28}$	0	0

Joint Density Function (Continuous)

- The function $f(x, y)$ is a **joint density** (joint pdf) of continuous random variables X and Y if:
 1. $f(x, y) \geq 0$
 2. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dx dy = 1$
 3. For any region A :

$$P((X, Y) \in A) = \iint_A f(x, y) dx dy$$

Example (Joint pdf)

- Let X be reaction time (s) and Y be temperature ($^{\circ}\text{F}$).
- Joint density:

$$f(x, y) = \begin{cases} 4xy, & 0 < x < 1, 0 < y < 1, \\ 0, & \text{elsewhere.} \end{cases}$$

- Find $P(0 \leq X \leq \frac{1}{2} \text{ and } \frac{1}{4} \leq Y \leq \frac{1}{2})$.

Example (Solution)

Problem 3.61: Find k

- Find k so that $f(x, y)$ is a valid joint pdf:

$$f(x, y) = \begin{cases} kxy, & 0 < x < 1, 0 < y < 1, x + y \leq 1, \\ 0, & \text{elsewhere.} \end{cases}$$

Marginal Distributions

- The marginal distributions of X alone and Y alone are:

- Discrete:

$$g(x) = \sum_y f(x, y), \quad h(y) = \sum_x f(x, y)$$

- Continuous:

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

Marginal Distributions: Discrete Case Example (Figure 2)

$f(x, y)$	$x = 0$	$x = 1$	$x = 2$	Row total ($h(y)$)
$y = 0$	$\frac{3}{28}$	$\frac{9}{28}$	$\frac{3}{28}$	$\frac{15}{28}$
$y = 1$	$\frac{3}{14}$	$\frac{3}{14}$	0	$\frac{3}{7}$
$y = 2$	$\frac{1}{28}$	0	0	$\frac{1}{28}$
Col total ($g(x)$)	$\frac{5}{14}$	$\frac{15}{28}$	$\frac{3}{28}$	1

- Marginal pmf of X :
- Marginal pmf of Y :

Marginal Distributions: Discrete Case Example (Insurance Table)

$f(x, y)$	$y = 0$	$y = 100$	$y = 200$
$x = 100$	0.20	0.10	0.20
$x = 250$	0.05	0.15	0.30

- Marginal pmf of X :
- Marginal pmf of Y :

Marginal Distributions: Continuous Case Example

$$f(x, y) = \begin{cases} 4xy, & 0 < x < 1, 0 < y < 1, \\ 0, & \text{elsewhere.} \end{cases}$$

- Marginal pdf of X :
- Marginal pdf of Y :

Independence

- Let X and Y have joint distribution $f(x, y)$ and marginals $g(x)$ and $h(y)$.
- X and Y are **statistically independent** iff

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

for all (x, y) in their range.

Independence: Discrete Case (Show NOT independent)

$$g(100) = 0.5, \ g(250) = 0.5, \quad h(0) = 0.25, \ h(100) = 0.25, \ h(200) = 0.5.$$

Independence: Continuous Case

$$f(x, y) = \begin{cases} 4xy, & 0 < x < 1, 0 < y < 1, \\ 0, & \text{elsewhere.} \end{cases}$$

Statistical Independence (Multiple Variables)

- Let X_1, X_2, \dots, X_n have joint distribution $f(x_1, \dots, x_n)$ and marginals $f_1(x_1), f_2(x_2), \dots, f_n(x_n)$.
- They are **mutually independent** iff

$$f(x_1, x_2, \dots, x_n) = f_1(x_1)f_2(x_2) \cdots f_n(x_n)$$

for all (x_1, \dots, x_n) in their range.