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.

$P(X \leq 2)$

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.

for continuous, ~~PX~~ $P(a < X < b) \geq 0$

$P(X \leq a) = P(X < a)$

Examples: Discrete or Continuous?

Are the following random variables discrete or continuous?

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

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.

If discrete, probability mass function

continuous, probability density function

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	1/4	2
E_2	H	T	HT	1/4	1
E_3	T	H	TH	1/4	1
E_4	T	T	TT	1/4	0

X	$p(x)$
0	1/4
1	1/2
2	1/4

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

$P(X \leq 0) = .13 + .15 + .17 + .2 = .65$

$P(-3 \leq X \leq 1) = .15 + .17 + .2 + .15 = .67$

$P(X = 2) = 0$

Marbles Example

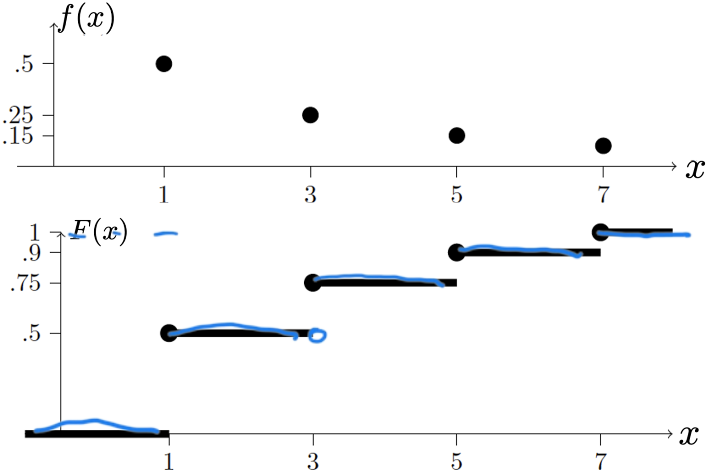
A box contains 8 red marbles and 3 green marbles. Six marbles are drawn at random without replacement. (How many outcomes result in all 3 green marbles being drawn?) Find the probability distribution for the number of green marbles drawn.

$$P(X=3) = \frac{\binom{3}{3}\binom{8}{3}}{\binom{11}{6}}$$

$x = 0, 1, 2, 3$

$$P(X=x) = \frac{\binom{3}{x}\binom{8}{6-x}}{\binom{11}{6}}$$

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

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

$$\lim_{x \to \infty} F(x) = 1$$

$$0 \leq F(x) \leq 1$$

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

$$F(x) = P(X \le x) = \begin{cases} 0, & x < -8 \\ 0.13, & -8 \le x < -3 \\ 0.28, & -3 \le x < -1 \end{cases}$$

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 $P(X \neq 0)$

$$F(x) = \begin{cases} 0, & x < 0 \\ 0.4, & 0 \leq x < 1 \\ 0.5, & 1 \leq x < 2 \\ 0.6, & 2 \leq x < 3 \\ 0.7, & 3 \leq x < 4 \\ 1, & x \geq 4 \end{cases}$$

$P(X \leq 2) = 0.6$

$P(X < 2) = 0.5$

$P(2 \leq X \leq 4) = 0.5$

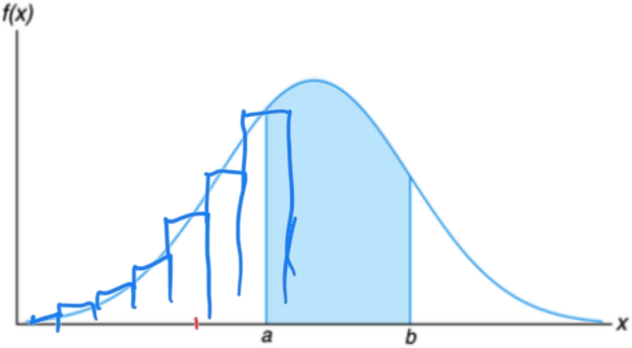
$P(X \neq 0) = 0.6$

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. $\displaystyle\int_{-\infty}^{\infty} f(x)\, dx = 1$
3. $\displaystyle P(a < X < b) = \int_a^b f(x)\, dx$

$f(x)$ has units of $\frac{1}{x}$ $\left(\frac{1}{cm} \text{ or } \frac{1}{in}\right)$

The Cumulative Distribution Function (Continuous)

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

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

$F(-\infty) = 0$

$F(\infty) = 1$

Example

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

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

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

$$F(x) = \begin{cases} 0 & x < 0 \\ \int_0^x \left(\frac{1}{8}x + \frac{3}{8}\right) dy = \frac{1}{16}x^2 + \frac{3}{8}x, & 0 \le x \le 2 \\ 1, & x \ge 2 \end{cases}$$

Example

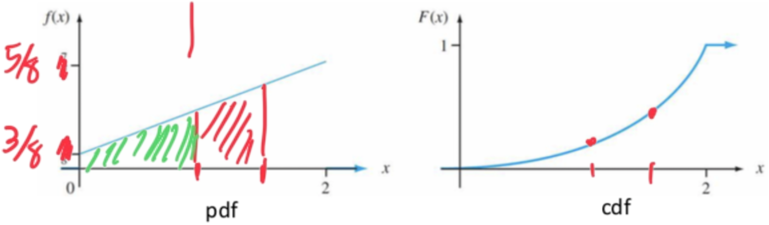


Figure: pdf and CDF

$P(1<x<1.5)$
$= F(1.5)-F(1)$
$= .266$

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

$P(x>1)=\int_1^2 f(x)\,dx = 1-F(1) = 1-(\frac{1}{16}+\frac{3}{16}) = .56$

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 $P(0.3 < X < 0.6)$

$$\int_0^1 \frac{k}{\sqrt{x}}\,dx = k\left[2\sqrt{x}\right]_0^1 = 2k = 1$$

$$\boxed{k = 1/2}$$

$$F(x) = \begin{cases} 0, & x \le 0 \\ \sqrt{x}, & 0 < x < 1 \\ 1, & x \ge 1 \end{cases}$$

$$P(0.3 < X < .6) = \sqrt{.6} - \sqrt{.3} = .227$$

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)

$y = f(x_1, x_2)$

$f(x_1, x_2, x_3)$

- 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

$P(X=250, Y=200)$

$= f(250, 200) = .3$

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. $\displaystyle\sum_x \sum_y f(x, y) = 1$
  3. $P(X = x,\ Y = y) = f(x, y)$
- For any region A in the xy-plane:

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

Example 3.14

- Two ballpoint pens are selected at random from a box with:

$$3 \text{ blue}, \; 2 \text{ red}, \; 3 \text{ green} \quad (8 \text{ total}).$$

- X = number of blue pens selected, Y = number of red pens selected.
- Find the joint pmf $f(x, y)$.

Total # of choices is $\binom{8}{2}$

$$f(x,y) = \frac{\binom{3}{x}\binom{2}{y}\binom{3}{2-x-y}}{\binom{8}{2}}; \; x \geq 0, \; y \geq 0, \; 2-x-y \geq 0$$

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:

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

Example (Joint pdf)

- Let X be reaction time (s) and Y be temperature (°F).
- Joint density:

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

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

$$P(0 \le X \le 1/2, 1/4 \le Y \le 1/2) = \int_0^{1/2} \int_{1/4}^{1/2} 4xy \, dy \, dx$$

$$= \int_0^{1/2} \left[2xy^2\right]_{1/4}^{1/2} dx$$

$$\left[x^2\right]_0^{1/2} \left[y^2\right]_{1/4}^{1/2}$$

Example (Solution)

$$\frac{1}{4} \times \left(\frac{1}{4} - \frac{1}{16}\right)$$

$$= \frac{3}{64}$$

Problem 3.61: Find k

$$\boxed{k=24}$$

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

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

$$1 = \int_0^1 \int_0^{1-x} kxy\,dy\,dx$$

$$= k\int_0^1 x\left[\frac{y^2}{2}\right]_0^{1-x} dx$$

$$= k\int_0^1 x\,\frac{(1-x)^2}{2}\,dx = \frac{k}{2}\int_0^1 (x-2x^2+x^3)\,dx$$

$$= \frac{k}{2}\left[\frac{x^2}{2}-\frac{2x^3}{3}+\frac{x^4}{4}\right]_0^1 = \frac{k}{2}\left(\frac{1}{12}\right) = \frac{k}{24}$$

Marginal Distributions

- The marginal distributions of X alone and Y alone are:
  - Discrete:
$$g(x) = \sum_y f(x, y), \qquad h(y) = \sum_x f(x, y)$$
  - Continuous:
$$g(x) = \int_{-\infty}^{\infty} f(x, y)\, dy, \qquad h(y) = \int_{-\infty}^{\infty} f(x, y)\, dx$$

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: $g(0)=5/14,\ g(1)=\frac{15}{28},\ g(2)=3/28$
- Marginal pmf of Y: $h(0)=\frac{15}{28},\ h(1)=3/7,\ h(2)=1/28$

Marginal Distributions: Discrete Case Example (Insurance Table)

$f(x, y)$	$y = 0$	$y = 100$	$y = 200$	
$x = 100$	0.20	0.10	0.20	.5
$x = 250$	0.05	0.15	0.30	.5
	.25	.25	.5	

- Marginal pmf of X: $g(x)$
- Marginal pmf of Y: $h(y)$

$$f(x,y) \overset{?}{\neq} g(x)h(y)$$

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: $g(x) = \int_0^1 4xy\,dy = 2x,\ 0<x<1$
- Marginal pdf of Y: $h(y) = \int_0^1 4xy\,dx = 2y,\ 0<y<1$

$\int_0^1 2x\,dx = 1$

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)

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

not independent

Independence: Continuous Case

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

$g(x) = 2x, \quad h(y) = 2y$

$f(x,y) = 4xy = g(x) \times h(y)$

X and Y are independent

Statistical Independence (Multiple Variables)

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

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

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