Chapter 5 – Random Variables and Discrete Probability Distributions

Section 5.1 – Random Variables

Random Variable – a rule that assigns a unique numerical value to each outcome in a sample space (typically named X , Y , etc)

Example

Experiment – Suppose that we flip a coin 3 times and record the sequence of tosses

Sample space – S = {HHH, HHT, HTH, HTT, THH, THT, TTH, TTT}

Random Variable – Suppose that we let X = the number of heads in an outcome

Experimental Outcome		Numerical Value of Random Variable
HHH		
HHT		
HTH	$\xrightarrow{X}$	
HTT		
THH		
THT		
TTH		
TTT		

Our focus will shift from looking at the experimental outcomes (e.g. HHH, HHT, etc) to looking at the values of X (e.g. 0, 1, 2, etc)

For example, in Chapter 4 we used

Outcome	HHH	HHT	HTH	HTT	THH	THT	TTH	TTT
Probability	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$

Now in Chapter 5 we'll use

x	0	1	2	3
Probability				

2 Types of Random Variables – A random variable is …

- … **discrete** if its set of possible values is a finite set or a countably infinite set (i.e. an infinite sequence with a 1st value, 2nd value, etc)
  (usually associated with counting, e.g. "the number of …")
- … **continuous** if its set of possible values is an infinite set that forms an interval on the number line (usually associated with measuring)

Note – These definitions are similar to those for discrete and continuous data (See Chapter 2)

Example (Exercise 5.11, p. 192) – Classify each random variable as discrete or continuous.

(a) The number of people requesting vegetarian meals on a flight from New York to London.

(b) The exact thickness (in millimeters) of a paper towel.

(c) The time it takes a driver to react after the car in front stops suddenly.

(d) The number of escapees in the next prison breakout.

Section 5.2 – Probability Distributions for Discrete Random Variables

Example – Let the discrete random variable X = the roll of an unbalanced or unfair or rigged die. Suppose the **probability distribution** (or **probability mass function**) of X is given by the following table

x	1	2	3	4	5	6
$p(x)$	0.40	0.20	0.12	0.10	0.14	?

Notation

- $p(x)$ = the amount of probability assigned to the specific value x
  $= P(X = x)$
- e.g. $p(2)$ = the amount of probability assigned to the specific value 2
  $= P(X = 2)$

(a) Find

(i) $p(6)$

(ii) $P(X = 1)$

(iii) The probability that the roll is at least 2.

(iv) The probability that the roll is at most 3.

(v) $P(1 < X \leq 4)$

(b) Draw a **probability histogram**.

(c) Suppose three people each roll the die. What is the probability that exactly two of them roll a four? (Hint – Use this for Exercise 5.60(c).)

(d) Suppose the die is rolled and $X > 2$. Find the chance that $X \geq 5$. (Hint – Use this for Exercise 5.34(e).)

Section 5.3 – Mean, Variance, and Standard Deviation for a Discrete R.V.

Important summaries of a data set are

- Sample mean = $\bar{x}$ (This indicates the center of the data)
- Sample standard deviation = s (This indicates the amount of spread among the data)

Likewise, a random variable or a probability distribution has similar summaries.

Definition – Let X be a discrete random variable with probability mass function $p(x)$.

- The **mean value** (μ) or the **expected value** ($E(X)$) of X is

Symbol		Calculation
μ or $E(X)$	=	$\sum_{All\ x} [x \cdot p(x)]$

- The **variance** of X is

Symbol		Calculation	
σ^2 or $Var(X)$	=	$\sum_{All\ x} [(x-\mu)^2 \cdot p(x)]$	(Definition)
		$E(X^2) - \mu^2$	(Short Cut)

- The **standard deviation** of X is

$$\sigma = \sqrt{\sigma^2}$$

Notes

- The mean, μ, gives the value of X that we would expect to see, on average. This indicates the center of a probability histogram.

- The standard deviation, σ, measures the amount of spread among the values of X or the amount of spread exhibited by the probability histogram. If σ is …
  - … large, then the values of X and/or the probability histogram has more spread
  - … small, then the values of X and/or the probability histogram has less spread

Example – Let the discrete random variable X = the roll of an unbalanced or unfair or rigged die. Suppose the probability distribution of X is given by the following table

x	1	2	3	4	5	6
$p(x)$	0.40	0.20	0.12	0.10	0.14	0.04

(a) Find the mean value (i.e. the expected value) of X .

(b) Find $E(X^2)$.

(c) Find the variance of X .

(d) Find the standard deviation of X.

Section 5.4 – The Binomial Distribution

Binomial Experiment – any experiment or situation that satisfies the following

- There is a known number of trials (denoted by n).
- Each trial results in a success or a failure.
- $P(\text{success})$ is the same for every trial (denoted by p).
- The trials are independent. That is, the outcome of one trial will not influence or affect the outcome of another trial.

Note – The values of n and p are called the **parameters** of the Binomial experiment.

Binomial Random Variable

- X = the total number of success observed during a Binomial experiment
- Possible values of X are 0, 1, 2, 3, … , n

Example – For each of the following decide whether or not the random variable is a Binomial random variable.

(a) A particular variety of seed has an 80% germination rate.
Let X = the number of seeds out of ten that germinate

(b) Five cards are drawn without replacement from a deck.
Let X = the number of red cards drawn

(c) You take a multiple choice quiz with five questions, each containing choices (a) to (d), by guessing.
Let X = the number of correct answers

Binomial Probability Mass Function

- $X \sim Bin(n, p)$ means that X is a Binomial random variable with parameters n and p
- $p(x)$ = the probability of obtaining exactly x successes among the n trials

$$= \binom{n}{x} \cdot p^x \cdot (1-p)^{n-x}$$

where

$$\binom{n}{x} = n \text{ choose } x = \frac{n!}{x!(n-x)!}$$

$$n! = n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot 2 \cdot 1$$

<u>Example</u> – Suppose that you take a multiple choice quiz with $n = 5$ questions, each containing choices (a) to (d), by guessing (so that $p = 0.25$). Let X = the number of correct answers. Here we know that X ~ Bin($n = 5$, $p = 0.25$).

(a) Find the probability of getting <u>exactly one</u> correct answer.

(b) Find the probability of getting <u>exactly two</u> correct answers.

(c) Find the probability of getting <u>at most one</u> correct answer.

(d) Find the probability of getting <u>at least four</u> correct answers.

Example – Suppose that $X \sim Bin(15, 0.40)$. Use the **tables of Binomial Distribution Cumulative Probabilities** on pages T-2 and T-3 (of your text book) or pages 7 and 8 (of your formula sheet) to find each of the following.

(a) $P(X \le 7)$

(b) $P(X > 5)$

(c) $P(X < 3)$

(d) $P(2 < X \le 6)$

(e) $P(5 \le X \le 9)$

(f) $P(3 < X < 7)$

(g) $P(X = 4)$

Binomial Mean, Variance, & Standard Deviation – If $X \sim Bin(n, p)$ then

- Mean: $\mu = n \cdot p$
- Variance: $\sigma^2 = n \cdot p \cdot (1 - p)$
- Standard Deviation: $\sigma = \sqrt{n \cdot p \cdot (1 - p)}$

Example – Suppose that you take a multiple choice quiz with n = 5 questions, each containing choices (a) to (d), by guessing (so that p = 0.25). Let X = the number of correct answers. Here we know that X ~ Bin(5,0.25).

(a) How many questions would you expect to get correct, i.e. what is the mean value of X ?

(b) Find the standard deviation of the number of correct answers, i.e. find σ.

Summary – Using the Binomial Tables

In class we discussed the probability mass function for the Binomial distribution

$$p(x) = \binom{n}{x} \cdot p^x \cdot (1-p)^{n-x} \qquad \text{for } x = 0, 1, 2, \ldots, n$$

The probability mass function is used to calculate the probability of exactly a successes, i.e. $P(X = a)$.

However, the Binomial tables are cumulative tables that provide the probability of at most a successes, i.e. $P(X \le a)$.

Below is a summary of how to use the tables to obtain various probabilities. You should study these until you understand them. They are not on the formula sheet and won't be provided at the exams.

- $P(X \le a)$ = table entry for a
- $P(X < a) = P(X \le a - 1)$ e.g. $P(X < 5) = P(X \le 4)$
- $P(X \ge a) = 1 - P(X \le a - 1)$ e.g. $P(X \ge 5) = 1 - P(X \le 4)$
- $P(X > a) = 1 - P(X \le a)$ e.g. $P(X > 5) = 1 - P(X \le 5)$
- $P(X = a) = P(X \le a) - P(X \le a - 1)$ e.g. $P(X = 5) = P(X \le 5) - P(X \le 4)$
- $P(a < X \le b) = P(X \le b) - P(X \le a)$
  e.g. $P(2 < X \le 5) = P(X \le 5) - P(X \le 2)$
- $P(a \le X \le b) = P(X \le b) - P(X \le a - 1)$
  e.g. $P(2 \le X \le 5) = P(X \le 5) - P(X \le 1)$
- $P(a \le X < b) = P(X \le b - 1) - P(X \le a - 1)$
  e.g. $P(2 \le X < 5) = P(X \le 4) - P(X \le 1)$
- $P(a < X < b) = P\,(X \le b - 1) - P(X \le a)$
  e.g. $P(2 < X < 5) = P(X \le 4) - P(X \le 2)$

Important note – These rules hold for discrete random variables but not necessarily for continuous random variables.