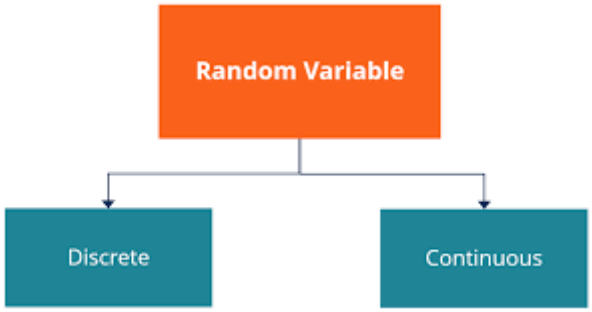
Chapter 4: Discrete Random Variables

(Discrete data are data that can be counted)

Vocabulary Review
Discrete data vs Continuous data
Independent events
Mutually exclusive events
With replacement vs Without replacement
Complement
Addition Rule for Mutually Exclusive events: P(A or B) = P(A) + P(B)

Introduction

- **Random variable** - a variable whose outcome is determined by chance. It represents a numerical quantity associated with each outcome of a probability distribution.



Upper case letters such as *X* or *Y* denote a random variable. Lower case letters like *x* or *y* denote the value of a random variable.

If *X* is a random variable, then *X* is written in words, and *x* is given as a number.

For example,

Let *X*=Number of sales calls a salesperson makes in one day. (a discrete random variable)
Then an example for x could be *x* = 24 calls in one day

Let *X* = Hours spent on sales calls in one day. (a continuous random variable)
Then an example for x could be *x* = 6.357 hours in one day

Let *X* = the number of heads you get when you toss three fair coins. The sample space (all possible outcomes) for the toss of three fair coins is

TTT THH HTH HHT HTT THT TTH HHH

Then, *x* = 0,1,2,3. *X* is in words and *x* is a number. Notice that for this example, the *x* values are countable outcomes. Because you can count the possible values that *X* can take on and the outcomes are random (the *x* values 0,1,2,3), *X* is a discrete random variable.

4.1: Probability Distribution for a Discrete Random Variable

- **Probability distribution** - a description that gives the probability for each value of a random variable. It provides the possible values of the random variable *X* and their corresponding probabilities. A probability distribution can be in the form of a table, graph, or mathematical formula.

x	$P(x)$
0	0.06
1	0.58
2	0.22
3	0.10
4	0.03
5	0.01

- **Probability mass function (PMF)** - a probability distribution for a **discrete** random variable
  - Notated as $f(x)$ where $f(x) = P(X = x)$
  - Properties
    - $0 \le f(x) \le 1$ (each probability is between 0 and 1, inclusive)
    - $f(x) = 0$ for every x not in the sample space
    - The sum of all the probabilities must be one

Example 1: Can the following tables be a probability distribution? State why or why not

x	$P(x)$
0	0.16
1	0.18
2	0.22
3	0.10
4	0.30
5	−0.01

x	$P(x)$
0	0.16
1	0.18
2	0.22
3	0.10
4	0.30
5	0.01

x	$P(x)$
0	0.06
1	0.58
2	0.22
3	0.10
4	0.03
5	0.01

Example 2: The following is a probability distribution. Determine the value of the missing probability.

x	*P(x)*
0	0.45
1	?
2	0.15
3	0.2

Inequalities Review for Sec 4.2

Inequality Symbols	
$\neq$	not equal
$<$	less than
$\leq$	less than or equal to
$>$	greater than
$\geq$	greater than or equal to

4.2: Mean or Expected Value and Standard Deviation

Expected Value (**Mean of the probability distribution of a discrete random variable**)

- Weighted average of all possible values over the sample space
- $\mu = \sum(x * f(x))$

Variance of a discrete random variable

- Weighted average of squared deviation about the mean
- $\sigma^2 = \sum[x^2 * f(x)] - \mu^2$

Standard deviation of a discrete random variable

- $\sigma = \sqrt{\sum[x^2 * f(x)] - \mu^2}$

Example 3: A baker is deciding how many batches of muffins to sell in his bakery. Through observation, the baker has established a probability distribution.

x	$f(x)$
1	0.15
2	0.35
3	0.40
4	0.10

(a) What is the probability of selling at least 3 batches?

(b) What is the probability of selling 1 or 4 batches of muffins?

(c) What is the probability of selling exactly 2 batches of muffins?

(d) What is the expected number of batches sold in a day? (Include units)

(e) What is the standard deviation of batches sold in a day? (include units)
Round to three decimal places.

Example 4: Suppose Nancy has classes three days a week. She attends classes three days a week 80% of the time, two days 15% of the time, one day 4% of the time, and no days 1% of the time.

(a) Find the expected value for class attendance *X*.

X	*P(x)*
0	0.01
1	0.04
2	0.15
3	0.80

(b) Find the standard deviation. Round to three decimal places

4.3: Binomial Distribution

- A Bernoulli trial is defined as a random experiment where only one of two mutually exclusive outcomes can occur.

- A **binomial distribution** is a sequence of Bernoulli trials such that we observe a fixed number of independent trials and the probability of success remains constant from trial to trial.

- If the random variable X is a binomial random variable, we denote it as $\boldsymbol{X \sim bin(n, p)}$ where *n* is the possible number of trials and *p* is the probability of a success in a single trial, and *x* represents the number of successes

Notation for Binomial Experiments

Symbol	*Description*
n	The number of times a trial is repeated.
p = P(S)	The probability of success in a single trial.
q = P(F)	The probability of failure in a single trial (q = 1 – p)
x	The random variable represents a count of the number of successes in n trials: x = 0, 1, 2, 3, . . . n.

Properties of a Binomial Distribution

- Two mutually exclusive outcomes (success and failure)
- Fixed number of independent trials
- P(success) remains constant from trial to trial

Example 5. Determine whether the following procedures results in a binomial distribution. If not, give an explanation.

(a) Consider the experiment where three marbles are drawn without replacement from a bag containing 20 red and 40 blue marbles, and the number of red marbles drawn are recorded.

(b) A fair-sided die is rolled ten times and the number of 6's is recorded.

(c) Choosing 8 marbles from a box of 40 marbles (20 purple, 12 red, and 8 green) one at a time with replacement keeping track of the colors of the marbles chosen.

(d) Choosing 10 marbles from a box of 40 (20 purple, 12 red, and 8 green) one at a time with replacement keeping track of the red marbles chosen.

(e) Choosing marbles from a box of 40 one at a time with replacement, until a red one is chosen.

Example 6.

Mr. Smith's history class has 16 male students and 13 female students. Ten different students from this class are selected *without* replacement, and their current GPAs are recorded.

- ○ Not binomial; there is not a fixed number of trials
- ○ Not binomial; there are more than two outcomes for each trial
- ○ Not binomial; the trials are not independent
- ○ Not binomial; for **more than one** of the reasons given in the above answer choices
- ○ This procedure results in a binomial distribution

Probability of a Binomial Distribution

Binomial probability formula: $f(x) = \binom{n}{x} p^x q^{n-x},\ where\ x = 0,1, \ldots n$

Mean of a binomial random variable: np

Variance of a binomial random variable: npq

Standard deviation of a binomial random variable: ?

Example 7. American Airlines flights from Dallas to Chicago are on time 79% of the time. Suppose 6 flights are randomly selected, and the number of on-time flights is recorded.

(a) Does this result in a binomial distribution? If so, identify the values of n, p, and q.

(b) Find the probability of **any** one flight being on time.

(c) Find the probability that exactly one flight is on time. Set up the calculation, but do not solve.

(d) Find the probability that exactly 3 flights are on time. Set it up and use a binomial calculator to solve.

(e) Find the probability that at most 3 flights are on time.

(f) Find the probability that at least 4 flights are on time. (hint: use the complement)

(g) Find the expected number of AA flights from Dallas to Chicago that will be on time and identify the correct units.

(h) Find the variance and identify the correct units.

(i) Find the standard deviation and identify the correct units.