Chapter 4: Discrete Probability Distributions

- In Chapter 2, we estimated population parameters such as μ, σ, using their sample analogues, $\bar{x}$, s, etc.
- These estimates were computed without making any assumptions about the data generating mechanism from which the data could have arisen.
- In this chapter, we study probability models for certain discrete data.
- We will then calculate population parameters such as μ and σ using the probability model.
- These notes discuss Sections 4-1 and 4-2 of the textbook.

Chapter 4: Discrete Probability Distributions

- A **random variable** is the outcome variable from an experiment whose outcome is governed by a chance mechanism.
- A **probability distribution** is a graph, a table or a formula that provides the probability of each value of the random variable.

Consider the experiment where you roll a six-sided fair die once and observe the outcome. The random variable of interest is

$$X = \text{outcome of the single roll of a fair die.}$$

The probability distribution is given by

x	1	2	3	4	5	6
$P(x)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

Here, $P(x)$ means $P(X = x)$.

- A couple decides to have four children.
- Consider the random variable X, where
  $X =$ number of girls in a family of four children.
- We will derive the probability distribution of X later. It is given by

x	0	1	2	3	4
$P(x)$	$\frac{1}{16}$	$\frac{1}{4}$	$\frac{3}{8}$	$\frac{1}{4}$	$\frac{1}{16}$

- Observe that all probabilities sums to one.
- What is probability that the couple will have at least two girls?
  $P(X \geq 2) = P(2) + P(3) + P(4) = 11/16$. Or, use
  $P(X \geq 2) = 1 - P(X < 2) = 1 - P(X \leq 1) =$
  $1 - [P(0) + P(1)] = 1 - [1/16 + 1/4] = 11/16$.

Chapter 4:

Requirements of a Probability Distribution

- In this chapter, we study the probability distributions of certain discrete random variables, i.e. the random variables take on either a finite number of values or are countably infinite.

Requirements of a Probability Distribution

(i) $0 \leq P(x) \leq 1$, for all x in the sample space.

(ii) $\sum_{\text{all } x} P(x) = 1$

- Probability Distribution: Probabilities are assigned based on a model.
- Relative Frequency Distribution: Probabilities are assigned based on repeating the experiment many times (i.e. data).

An experiment involves groups of four seedlings grown under controlled conditions. The probability distribution of X, the number of seedlings in a groups that are classified as diseased is

x	0	1	2	3	4
$P(x)$	?	0.113	0.057	0.009	0.002

An experiment involves groups of four seedlings grown under controlled conditions. The probability distribution of X, the number of seedlings in a groups that are classified as diseased is

x	0	1	2	3	4
$P(x)$	?	0.113	0.057	0.009	0.002

$P(0) = 1 - (P(1) + P(2) + P(3) + P(4)) = 0.819.$

The probability that monkeys will have a positive reaction during a trial of an experiment is $1/3$. Define the random variable X to be the number of trials performed until a positive reaction is observed. Assume independence of trials.

(a) Find the probability distribution of X.

(b) What is $P(X \geq 3)$?

(a) Let's calculate $P(X = x)$ for $x = 1, 2, 3$ for starters. Observe by definition, $P(X = 1) = 1/3$. Next, $P(X = 2) =$ $P(\text{first trial is a neg. reac'n and second trial is a pos. reac'n}) =$ $P(\text{first trial is negative}) \times P(\text{second is positive}) = 2/3 \times 1/3$ by independence of trials (given fact from question). By similar reasoning, $P(X = 3) = (2/3)^2 \times 1/3$. This pattern suggests

$$P(X = x) = \left(\frac{2}{3}\right)^{x-1} \frac{1}{3}, \text{ for } x = 1, 2, \ldots$$

(b) $P(X \geq 3) = 1 - P(X < 3) = 1 - [P(1) + P(2)] =$ $1 - [1/3 + 2/9] = 4/9$.

- We can also talk about measures of centre and spread for random variables because the probabilities given by the model can be viewed as long-range relative frequencies.
- Consider the following data:

$$0, \quad 2, \quad 2, \quad 1, \quad 2, \quad 3, \quad 0, \quad 1, \quad 2, \quad 1.$$

Then,

$$\begin{aligned}
\bar{x} &= \frac{0+2+\ldots+1}{10} \\
&= 0.\left(\frac{2}{10}\right)+1.\left(\frac{3}{10}\right)+2.\left(\frac{4}{10}\right)+3.\left(\frac{1}{10}\right) \\
&= \sum \text{value} \times \text{relative frequency}
\end{aligned}$$

Mean

$$\mu = E(X) = \sum[x \cdot P(x)]$$

Variance

$$\begin{aligned} \sigma^2 = Var(X) &= \sum[(x-\mu)^2 \cdot P(x)] \\ &= \sum[x^2 \cdot P(x)] - \mu^2 \quad \text{short-cut formula} \end{aligned}$$

Standard Deviation

$$\sigma = \sqrt{\sum[(x-\mu)^2 \cdot P(x)]}.$$

Chapter 4: Section 4-2

1. $E(X)$ is also referred to as the *expectation*, or the *expected value* of the random variable X.
2. Best to use short-cut formula for standard deviation.
3. The mean, variance, and standard deviation of a random variable X, are population quantities. Hence, the use of Greek letters μ, σ^2, and σ.

Example 2 Revisited

Find the mean and the standard deviation of X, the number girls in a family of four children. The probability distribution of X is

x	0	1	2	3	4
$P(x)$	$\frac{1}{16}$	$\frac{1}{4}$	$\frac{3}{8}$	$\frac{1}{4}$	$\frac{1}{16}$

Example 2 Revisited

Find the mean and the standard deviation of X, the number girls in a family of four children. The probability distribution of X is

x	0	1	2	3	4
$P(x)$	$\frac{1}{16}$	$\frac{1}{4}$	$\frac{3}{8}$	$\frac{1}{4}$	$\frac{1}{16}$

(a) Mean:
$$\mu = 0 \cdot \left(\frac{1}{16}\right) + 1 \cdot \left(\frac{1}{4}\right) + 2 \cdot \left(\frac{3}{8}\right) + 3 \cdot \left(\frac{1}{4}\right) + 4 \cdot \left(\frac{1}{16}\right) = 2$$

(b) Variance:
$$\sigma^2 = \left[0^2 \cdot \frac{1}{16} + 1^2 \cdot \left(\frac{1}{4}\right) + 2^2 \cdot \left(\frac{3}{8}\right) + 3^2 \left(\cdot \frac{1}{4}\right) + 4^2 \cdot \left(\frac{1}{16}\right)\right] - 2^2 = 1.$$

(c) Standard Deviation: $\sigma = \sqrt{1} = 1.$