Chapter 5 Discrete Probability Distributions

5.1 Probability Distributions

Objectives

In this lesson, we will learn how to

- identify random variables and probability distributions;
- distinguish between discrete and continuous random variables;
- check whether a table is a valid probability distribution;
- make and interpret probability histograms;
- find the mean, variance, and standard deviation of a probability distribution;
- identify significantly low or significantly high values;
- compute and interpret expected value.

1 Key Concept

This section introduces two important ideas: random variables and probability distributions.

A probability distribution allows us to describe the possible outcomes of a random process and the probability assigned to each outcome.

Main Ideas

- A random variable assigns a numerical value to each outcome of a chance process.
- A probability distribution lists the possible values of the random variable and their probabilities.

- We can use a probability distribution to find the mean, variance, and standard deviation.
- We can also decide whether an outcome is unusually low or unusually high.

2 Random Variables

Random Variable

A random variable is a variable, usually represented by x, that has a single numerical value determined by chance for each outcome of a procedure.

For example, suppose a family has two children. Let

$$x = \text{the number of females among the two children.}$$

Then x is a random variable because its value depends on chance. The possible values are

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

3 Probability Distributions

Probability Distribution

A probability distribution gives the probability for each value of a random variable. It is often written as a table, formula, or graph.

For two births, assuming male and female births are equally likely, the probability distribution for the number of females is:

x: number of females in two births	$P(x)$
0	0.25
1	0.50
2	0.25

This table tells us:

- $P(0) = 0.25$, so there is a 25% chance of no females.
- $P(1) = 0.50$, so there is a 50% chance of one female.
- $P(2) = 0.25$, so there is a 25% chance of two females.

4 Discrete and Continuous Random Variables

Discrete Random Variable

A discrete random variable has values that are finite or countable.

Examples:

- number of females in two births;
- number of heads in five coin tosses;
- number of students absent from class today.

Continuous Random Variable

A continuous random variable has infinitely many possible values that are not countable one by one. These values usually come from measurements.

Examples:

- body temperature;
- height;
- weight;
- time required to complete an exam.

5 Requirements for a Probability Distribution

Every probability distribution must satisfy three requirements.

Requirements for a Probability Distribution

1. The variable x must be a numerical random variable.
2. The sum of all probabilities must equal 1:

$$\sum P(x) = 1.$$

   Small differences such as 0.999 or 1.001 may occur because of rounding.

3. Each probability must be between 0 and 1:

$$0 \leq P(x) \leq 1.$$

Example 1: For two births, let x be the number of females. Determine whether the following table is a probability distribution.

x: number of females in two births	$P(x)$
0	0.25
1	0.50
2	0.25

Solution

We check the three requirements.

1. x is numerical because $x = 0, 1, 2$.
2. The probabilities add to 1:

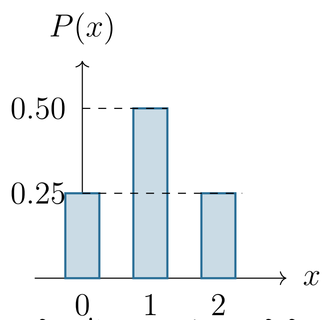
$$0.25 + 0.50 + 0.25 = 1.$$

3. Each probability is between 0 and 1.

Therefore, the table is a probability distribution.

6 Probability Histograms

A probability histogram is similar to a relative frequency histogram, but the vertical scale shows probabilities.



Probability histogram for the number of females in two births

The bars show the probability for each possible value of x.

7 Probability Distribution from a Formula

A probability distribution can also be written as a formula.

For the number of females in two births, we can use

$$P(x) = \frac{2!}{x!(2-x)!}\left(\frac{1}{2}\right)^2, \qquad x = 0, 1, 2.$$

Then

$$P(0) = 0.25, \qquad P(1) = 0.50, \qquad P(2) = 0.25.$$

These match the probabilities in the table.

8 Checking Whether a Table Is a Probability Distribution

Example 2: The table lists countries and the proportion of unlicensed software in each country. Does the table describe a probability distribution?

Country	Proportion of Unlicensed Software
United States	0.17
China	0.70
India	0.58
Russia	0.64
Total	2.09

Solution

This table is not a probability distribution.
It violates the requirements because:

- the values are countries, which are categorical, not numerical random-variable values;
- the total is 2.09, but a probability distribution must have

$$\sum P(x) = 1.$$

Therefore, the table does not describe a probability distribution.

9 Parameters of a Probability Distribution

For a probability distribution, we are describing a population, not a sample. Therefore, the mean, variance, and standard deviation are parameters.

Mean of a Discrete Probability Distribution

$$\mu = \sum xP(x)$$

Variance of a Discrete Probability Distribution

$$\sigma^2 = \sum (x-\mu)^2 P(x)$$

An equivalent shortcut formula is

$$\sigma^2 = \sum x^2 P(x) - \mu^2.$$

Standard Deviation of a Discrete Probability Distribution

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

10 Expected Value

Expected Value

The expected value of a discrete random variable is the mean value of the outcomes.

$$E(x) = \mu = \sum xP(x)$$

Expected value is a long-run average. It does not have to be one of the possible values of the random variable.

11 Finding Mean, Variance, and Standard Deviation

Example 3: The table describes the probability distribution for the number of females in two births. Find the mean, variance, and standard deviation.

x: number of females in two births	$P(x)$
0	0.25
1	0.50
2	0.25

Solution

We create a table for the calculations.

$\mathbf{x}$	$\mathbf{P(x)}$	$\mathbf{xP(x)}$	$\mathbf{x^2P(x)}$
0	0.25	$0(0.25) = 0$	$0^2(0.25) = 0$
1	0.50	$1(0.50) = 0.50$	$1^2(0.50) = 0.50$
2	0.25	$2(0.25) = 0.50$	$2^2(0.25) = 1.00$
Total	1.00	1.00	1.50

The mean is

$$\mu = \sum xP(x) = 1.00.$$

The variance is

$$\sigma^2 = \sum x^2P(x) - \mu^2 = 1.50 - (1.00)^2 = 0.50.$$

The standard deviation is

$$\sigma = \sqrt{0.50} \approx 0.7.$$

Therefore,

$$\mu = 1.0, \qquad \sigma^2 = 0.5, \qquad \sigma \approx 0.7.$$

Interpretation: In the long run, we expect an average of about 1.0 female in two births.

12 Identifying Significant Results with the Range Rule of Thumb

Range Rule of Thumb

- Significantly low values are

$$\mu - 2\sigma \text{ or lower.}$$

- Significantly high values are

$$\mu + 2\sigma \text{ or higher.}$$

- Values between $\mu - 2\sigma$ and $\mu + 2\sigma$ are not considered significant by this rule.

Example 4: For two births, the mean number of females is

$$\mu = 1.0$$

and the standard deviation is

$$\sigma = 0.7.$$

Use the range rule of thumb to determine whether 2 females in 2 births is significantly high.

Solution

A value is significantly high if it is greater than or equal to

$$\mu + 2\sigma.$$

Substitute the values:

$$\mu + 2\sigma = 1.0 + 2(0.7) = 1.0 + 1.4 = 2.4.$$

Significantly high values are 2.4 or higher.

Since

$$2 < 2.4,$$

2 females in 2 births is not significantly high.

13 Identifying Significant Results with Probabilities

Significantly High

x successes among n trials is significantly high if

$$P(x \text{ or more}) \leq 0.05.$$

Significantly Low

x successes among n trials is significantly low if

$$P(x \text{ or fewer}) \leq 0.05.$$

The value 0.05 is common, but it is not absolutely rigid. Sometimes other values, such as 0.01, may be used.

14 Rare Event Rule for Inferential Statistics

Rare Event Rule

If, under a given assumption, the probability of a particular outcome is very small and the outcome occurs significantly less than or significantly greater than expected, we conclude that the assumption is probably not correct.

This idea is one of the foundations of hypothesis testing.

15 Expected Value in Gambling

Expected value helps us compare choices when outcomes involve gains and losses.

Expected Value with Payoffs

$$E(x) = \sum xP(x)$$

where x represents the payoff or net gain for each outcome.

Example 5: You have $\$5$ to place on a bet in a casino. Compare the expected values of the following two bets.

- Roulette: Bet $\$5$ on the number 7. The probability of losing $\$5$ is $\frac{37}{38}$, and the probability of a net gain of $\$175$ is $\frac{1}{38}$.
- Craps: Bet $\$5$ on the pass line. The probability of losing $\$5$ is $\frac{251}{495}$, and the probability of a net gain of $\$5$ is $\frac{244}{495}$.

Which bet is better in the sense of having the higher expected value?

Solution

For roulette:

$$E(x) = (-5)\left(\frac{37}{38}\right) + (175)\left(\frac{1}{38}\right).$$

$$E(x) = \frac{-185}{38} + \frac{175}{38} = \frac{-10}{38} \approx -0.26.$$

So the expected value for roulette is about

$$-\$0.26.$$

For craps:

$$E(x) = (-5)\left(\frac{251}{495}\right) + (5)\left(\frac{244}{495}\right).$$

$$E(x) = \frac{-1255}{495} + \frac{1220}{495} = \frac{-35}{495} \approx -0.07.$$

So the expected value for craps is about

$$-\$0.07.$$

Because losing about 7 cents is better than losing about 26 cents, the craps bet is better in the long run.

Summary

Key Ideas

- A random variable assigns a numerical value to each outcome of a chance process.
- A probability distribution gives the probability for each value of the random variable.
- A discrete random variable has finite or countable possible values.
- A continuous random variable has infinitely many possible values on a continuous scale.
- A probability distribution must satisfy $\sum P(x) = 1$ and $0 \leq P(x) \leq 1$.
- The mean of a probability distribution is

$$\mu = \sum xP(x).$$

- The variance and standard deviation are

$$\sigma^2 = \sum x^2P(x) - \mu^2, \qquad \sigma = \sqrt{\sigma^2}.$$

- Expected value is the long-run average outcome.
- The range rule of thumb and probability rules help identify unusually low or unusually high outcomes.