Discrete Probability Distribution

Binomial Probability Distribution

Chapter 5

Random variable
- has a single numerical value, determined by chance, for each outcome of a procedure.
- Random variables are usually denoted X or Y

Discrete
- takes on a countable number of values (i.e. there are gaps between values).

Continuous
- there are an infinite number of values the random variable can take, and they are densely packed together (i.e. there are no gaps between values).

Would the following random variable, $X$, be discrete or continuous?

$X$ = the number of sales at the drive-through during the lunch rush at the local fast food restaurant.

a) Continuous
b) Discrete

Would the following random variable, $X$, be discrete or continuous?

$X$ = the time required to run a marathon.

a) Continuous
b) Discrete

Would the following random variable, $X$, be discrete or continuous?

$X$ = the number of fans in a football stadium.

a) Continuous
b) Discrete

Would the following random variable, $X$, be discrete or continuous?

$X$ = the distance a car could drive with only one gallon of gas.

a) Continuous
b) Discrete
Random variables

- A probability distribution is a description of the chance a random variable has of taking on particular values. It is often displayed in a graph, table, or formula.
- A probability histogram is a display of a probability distribution of a discrete random variable. It is a histogram where each bar’s height represents the probability that the random variable takes on a particular value.

Example

An industrial psychologist administered a personality inventory test for passive-aggressive traits to 150 employees. Individuals were rated on a score from 1 to 5, where 1 was extremely passive and 5 extremely aggressive. A score of 3 indicated neither trait. The results are shown below:

<table>
<thead>
<tr>
<th>Score, X</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency</td>
<td>24</td>
<td>33</td>
<td>42</td>
<td>30</td>
<td>21</td>
</tr>
</tbody>
</table>

Discrete Random Variable (DRV)
Probability Distribution

Properties:

- Discrete probability distribution includes all the values of DRV;
- For any value x of DRV, 0 ≤ P(x) ≤ 1;
- The sum of probabilities of all the DRV values equals to 1;
- The values of DRV are mutually exclusive.

Example cont.

Construct the probability distribution:

<table>
<thead>
<tr>
<th>Score, X</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability P(X=x)</td>
<td>24/150 = 0.16</td>
<td>33/150 = 0.22</td>
<td>42/150 = 0.28</td>
<td>30/150 = 0.20</td>
<td>21/150 = 0.14</td>
</tr>
</tbody>
</table>

- Verify that it’s a probability distribution.
- For all values of X, 0 ≤ P(x) ≤ 1
- 0.09 + 0.36 + 0.35 + 0.13 + 0.05 + 0.02 = 1

Example cont.

- Graph the distribution.

Note: the area of each bar is equal to the probability of a particular outcome.

Example cont.

- What is the probability that a randomly selected worker got a score 3 or less?
  - P(X ≤ 3) = 0.16 + 0.22 + 0.28 = 0.66
- What is the probability that a randomly selected worker scored at least 4?
  - P(X ≥ 4) = 0.2 + 0.14 = 0.34
- What is the probability that a randomly selected worker did not score 5?
  - P(X ≠ 5) = 1 - 0.14 = 0.86
### The Mean and Standard Deviation of a Discrete Random Variable

- **The mean** or **expected value** of a discrete random variable is given by
  \[ \mu = \sum x P(x) = x_1 \cdot P(x_1) + x_2 \cdot P(x_2) + \ldots \]
- **The variance** of a discrete random variable is given by
  \[ \sigma^2 = \sum (x - \mu)^2 P(x) = (x_1 - \mu)^2 \cdot P(x_1) + (x_2 - \mu)^2 \cdot P(x_2) + \ldots \]
- **The standard deviation** is
  \[ \sigma = \sqrt{\sigma^2} \]

### Example cont.

**What is the mean score? What can you conclude?**

\[ \mu = \sum x P(x) = x_1 \cdot P(x_1) + x_2 \cdot P(x_2) + \ldots \]

\[ \mu = 1 \cdot 0.16 + 2 \cdot 0.22 + 3 \cdot 0.28 + 4 \cdot 0.2 + 5 \cdot 0.14 = 2.94 \]

*We can conclude that the mean personality trait is neither extremely passive nor extremely aggressive, but is slightly closer to passive.*

<table>
<thead>
<tr>
<th>Score, X</th>
<th>Probability P(X=x)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.16</td>
</tr>
<tr>
<td>2</td>
<td>0.22</td>
</tr>
<tr>
<td>3</td>
<td>0.28</td>
</tr>
<tr>
<td>4</td>
<td>0.2</td>
</tr>
<tr>
<td>5</td>
<td>0.14</td>
</tr>
</tbody>
</table>

### Example cont.

**Find the variance and standard deviation of the scores. What can you conclude?**

\[ \sigma^2 = \sum (x - \mu)^2 P(x) = (x_1 - \mu)^2 \cdot P(x_1) + (x_2 - \mu)^2 \cdot P(x_2) + \ldots \]

\[ \sigma^2 = (1 - 2.94)^2 \cdot 0.16 + (2 - 2.94)^2 \cdot 0.22 + (3 - 2.94)^2 \cdot 0.28 + (4 - 2.94)^2 \cdot 0.2 + (5 - 2.94)^2 \cdot 0.14 \]

\[ \sigma^2 = 1.616 \]

\[ \sigma = \sqrt{\sigma^2} = 1.616 \approx 1.3 \]

*Most of the data values differ from the mean by no more than 1.3 points.*

### A Special Discrete Random Variable: Binomial Random Variable

- **Binomial Random Variable (BRV) probability distribution is a special case of the DRV probability distribution, when there are only two outcomes:** Success and Failure;
- **BRV value is defined as the number of Successes in a given number of trials;**

**Conditions that must be met to consider the experiment as Binomial:**

- Every trial must have only two mutually exclusive outcomes: Success or Failure,
- The probability of Success and Failure must remain constant from trial to trial,
- The outcome of the trial is independent of the outcomes of the previous trials,
- There are a fixed number of trials.

### Binomial setting

A manufacturing company takes a sample of \( n = 100 \) bolts from their production line. \( X \) is the number of bolts that are found defective in the sample. It is known that the probability of a bolt being defective is 0.003. Does \( X \) have a binomial distribution?

- a) Yes.
- b) No, because there is not a fixed number of observations.
- c) No, because the observations are not all independent.
- d) No, because there are more than two possible outcomes for each observation.
- e) No, because the probability of success for each observation is not the same.

### Binomial setting

A survey-taker asks the age of each person in a random sample of 20 people. \( X \) is the age for the individuals. Does \( X \) have a binomial distribution?

- a) Yes.
- b) No, because there is not a fixed number of observations.
- c) No, because the observations are not all independent.
- d) No, because there are more than two possible outcomes for each observation.
Binomial setting
A survey-taker asks whether each person in a random sample of 20 college students is over the age of 21. According to university records, 35% of all college students are over 21 years old. Does X have a binomial distribution?

a) Yes.
b) No, because there is not a fixed number of observations.
c) No, because the observations are not all independent.
d) No, because there are more than two possible outcomes for each observation.
e) No, because the probability of success for each observation is not the same.

Binomial setting
A fair die is rolled and the number of dots on the top face is noted. X is the number of times we have to roll in order to have the face of the die show a 2. Does X have a binomial distribution?

a) Yes.
b) No, because there is not a fixed number of observations.
c) No, because the observations are not all independent.
d) No, because there are more than two possible outcomes for each observation.
e) No, because the probability of success for each observation is not the same.

Probability of x Successes in n trials

\[
P(X = x) = \binom{n}{x} \cdot p^x \cdot (1-p)^{n-x}
\]

- \(P\) = probability of x-successes in n-trials
- \(x\) = number of successes
- \(n\) = number of trials
- \(p\) = probability of a success in one trial

Example
For the number of trials \(n=5\), and for a probability of one success \(p=0.5\), the binomial probability distribution has the following shape:
Statistical Parameters of a Binomial Distribution

- Mean of the distribution:  \( \mu = np \)
- Standard Deviation:  \( \sigma = \sqrt{np(1 - p)} \)

\( n \) – number of trials;
\( p \) – probability of a success in one trial;

Binomial distribution

Suppose that for a randomly selected high school student who has taken a college entrance exam, the probability of scoring above a 650 is 0.30. A random sample of \( n = 9 \) students was selected. What are the mean \( \mu \) and standard deviation \( \sigma \) of the number of students in the sample who have scores above 650?

a)  \( \mu = 9(0.3) = 2.7, \sigma = 0.30 \)
b)  \( \mu = 3, \sigma = 9(0.3) \)
c)  \( \mu = 9(0.3) = 2.7, \sigma = 9(0.7)(0.3) \)
d)  \( \mu = 9(0.3) = 2.7, \sigma = \sqrt{9(0.3)(0.7)} \)