6.1: Discrete and Continuous Random Variables

THE BEST WAY TO MAKE THIS DECISION IS BY CALCULATING THE EXPECTED VALUE OF EACH POSSIBLE OUTCOME.

YOU MULTIPLY THE... MUST PRETEND TO BE DEAD.

I SENSE THAT WE'RE DONE HERE.

I HOPE THE DEAD SOMETIMES COVER THEIR EARS.
Section 6.1
Discrete & Continuous Random Variables

After this section, you should be able to...

✓ APPLY the concept of discrete random variables to a variety of statistical settings

✓ CALCULATE and INTERPRET the mean (expected value) of a discrete random variable

✓ CALCULATE and INTERPRET the standard deviation (and variance) of a discrete random variable

✓ DESCRIBE continuous random variables
Random Variables

• A random variable, usually written as X, is a variable whose possible values are numerical outcomes of a random phenomenon.
• There are two types of random variables, discrete random variables and continuous random variables.
Discrete Random Variables

• A discrete random variable is one which may take on only a countable number of distinct values such as 0, 1, 2, 3, 4,....

• Discrete random variables are usually (but not necessarily) counts.

• Examples:
  • number of children in a family
  • the Friday night attendance at a cinema
  • the number of patients a doctor sees in one day
  • the number of defective light bulbs in a box of ten
  • the number of “heads” flipped in 3 trials
Probability Distribution

The **probability distribution** of a discrete random variable is a list of probabilities associated with each of its possible values.

Consider tossing a fair coin 3 times. Define $X = \text{the number of heads obtained}$

<table>
<thead>
<tr>
<th>Value</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability</td>
<td>$1/8$</td>
<td>$3/8$</td>
<td>$3/8$</td>
<td>$1/8$</td>
</tr>
</tbody>
</table>
Rolling Dice: Probability Distribution

Roll your pair of dice 20 times, record the sum for each trial.
A discrete random variable $X$ takes a fixed set of possible values with gaps between. The probability distribution of a discrete random variable $X$ lists the values $x_i$ and their probabilities $p_i$:

<table>
<thead>
<tr>
<th>Value</th>
<th>$x_1$</th>
<th>$x_2$</th>
<th>$x_3$</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability</td>
<td>$p_1$</td>
<td>$p_2$</td>
<td>$p_3$</td>
<td>...</td>
</tr>
</tbody>
</table>

The probabilities $p_i$ must satisfy two requirements:

1. Every probability $p_i$ is a number between 0 and 1.
2. The sum of the probabilities is 1.

To find the probability of any event, add the probabilities $p_i$ of the particular values $x_i$ that make up the event.
Describing the (Probability) Distribution

When analyzing discrete random variables, we’ll follow the same strategy we used with quantitative data – describe the shape, center (mean), and spread (standard deviation), and identify any outliers.
Example: Babies’ Health at Birth

(a) Prove that the probability distribution for $X$ is legitimate.

(b) Apgar scores of 7 or higher indicate a healthy baby. What is $P(X \geq 7)$?

<table>
<thead>
<tr>
<th>Value:</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability:</td>
<td>0.001</td>
<td>0.006</td>
<td>0.007</td>
<td>0.008</td>
<td>0.012</td>
<td>0.020</td>
<td>0.038</td>
<td>0.099</td>
<td>0.319</td>
<td>0.437</td>
<td>0.053</td>
</tr>
</tbody>
</table>
Example: Babies’ Health at Birth

(a) Prove that the probability distribution for $X$ is legitimate.

(b) Apgar scores of 7 or higher indicate a healthy baby. What is $P(X \geq 7)$?

<table>
<thead>
<tr>
<th>Value:</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability:</td>
<td>0.001</td>
<td>0.006</td>
<td>0.007</td>
<td>0.008</td>
<td>0.012</td>
<td>0.020</td>
<td>0.038</td>
<td>0.099</td>
<td>0.319</td>
<td>0.437</td>
<td>0.053</td>
</tr>
</tbody>
</table>

(a) All probabilities are between 0 and 1 and the probabilities sum to 1. This is a legitimate probability distribution.

(b) $0.099 + 0.319 + 0.437 + 0.053 = P(X \geq 7) = .908$ We’d have a 91% chance of randomly choosing a healthy baby.
Example: Babies’ Health at Birth

C. Describe the probability distribution.

Value: 0 1 2 3 4 5 6 7 8 9 10
Probability: 0.001 0.006 0.007 0.008 0.012 0.020 0.038 0.099 0.319 0.437 0.053
## Mean of a Discrete Random Variable

The mean of any discrete random variable is an average of the possible outcomes, with each outcome weighted by its probability.

Suppose that $X$ is a discrete random variable whose probability distribution is

<table>
<thead>
<tr>
<th>Value</th>
<th>$x_1$</th>
<th>$x_2$</th>
<th>$x_3$</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability</td>
<td>$p_1$</td>
<td>$p_2$</td>
<td>$p_3$</td>
<td>...</td>
</tr>
</tbody>
</table>

To find the **mean (expected value)** of $X$, multiply each possible value by its probability, then add all the products:

$$
\mu_x = E(X) = x_1 p_1 + x_2 p_2 + x_3 p_3 + \ldots \\
= \sum x_i p_i
$$
Example: Apgar Scores – What’s Typical?

Consider the random variable $X = \text{Apgar Score}$

**Compute the mean of the random variable $X$ and interpret it in context.**

<table>
<thead>
<tr>
<th>Value</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability</td>
<td>0.001</td>
<td>0.006</td>
<td>0.007</td>
<td>0.008</td>
<td>0.012</td>
<td>0.020</td>
<td>0.038</td>
<td>0.099</td>
<td>0.319</td>
<td>0.437</td>
<td>0.053</td>
</tr>
</tbody>
</table>

$$\mu_x = E(X) = \sum x_i p_i$$
Example: Apgar Scores – What’s Typical?

Consider the random variable $X = \text{Apgar Score}$

**Compute the mean of the random variable** $X$ **and interpret it in context.**

<table>
<thead>
<tr>
<th>Value</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability:</td>
<td>0.001</td>
<td>0.006</td>
<td>0.007</td>
<td>0.008</td>
<td>0.012</td>
<td>0.020</td>
<td>0.038</td>
<td>0.099</td>
<td>0.319</td>
<td>0.437</td>
<td>0.053</td>
</tr>
</tbody>
</table>

$$\mu_x = E(X) = \sum x_i p_i$$

$$= (0)(0.001) + (1)(0.006) + (2)(0.007) + \ldots + (10)(0.053)$$

$$= 8.128$$

The mean Apgar score of a randomly selected newborn is 8.128. This is the long-term average Agar score of many, many randomly chosen babies.

**Note:** The expected value does not need to be a possible value of $X$ or an integer! It is a long-term average over many repetitions.
Analyzing Discrete Random Variables on the Calculator

1. Using one-variable statistics to calculate:
2. Enter “ascre” for X1 and “freqas” for frequency list.
Analyzing Discrete Random Variables on the Calculator
Standard Deviation of a Discrete Random Variable

The definition of the variance of a random variable is similar to the definition of the variance for a set of quantitative data. To get the standard deviation of a random variable, take the square root of the variance.

Suppose that $X$ is a discrete random variable whose probability distribution is

Value: $x_1$ $x_2$ $x_3$ ...
Probability: $p_1$ $p_2$ $p_3$ ...

and that $\mu_X$ is the mean of $X$. The variance of $X$ is

$$Var(X) = \sigma_X^2 = (x_1 - \mu_X)^2 p_1 + (x_2 - \mu_X)^2 p_2 + (x_3 - \mu_X)^2 p_3 + ...$$

$$= \sum (x_i - \mu_X)^2 p_i$$
Example: Apgar Scores – How Variable Are They?

Consider the random variable $X = \text{Apgar Score}$

**Compute the standard deviation of the random variable** $X$ **and interpret it in context.**

<table>
<thead>
<tr>
<th>Value:</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability:</td>
<td>0.001</td>
<td>0.006</td>
<td>0.007</td>
<td>0.008</td>
<td>0.012</td>
<td>0.020</td>
<td>0.038</td>
<td>0.099</td>
<td>0.319</td>
<td>0.437</td>
<td>0.053</td>
</tr>
</tbody>
</table>

$$\sigma_X^2 = \sum (x_i - \mu_X)^2 p_i$$

$$= (0 - 8.128)^2 (0.001) + (1 - 8.128)^2 (0.006) + ... + (10 - 8.128)^2 (0.053)$$

$$= 2.066$$

Variance

$$\sigma_X = \sqrt{2.066} = 1.437$$

The standard deviation of $X$ is 1.437. On average, a randomly selected baby’s Apgar score will differ from the mean 8.128 by about 1.4 units.
C. Describe the probability distribution.

C. The left-skewed shape of the distribution suggests a randomly selected newborn will have an Apgar score at the high end of the scale. While the range is from 0 to 10, there is a VERY small chance of getting a baby with a score of 5 or lower. There are no obvious outliers. The expected value of Apgar scores for a typical baby is 8.128.
Calculate the Mean (Expected Value)

<table>
<thead>
<tr>
<th>Value</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability</td>
<td>0.08</td>
<td>0.12</td>
<td>0.30</td>
<td>0.22</td>
<td>0.20</td>
<td>0.08</td>
</tr>
</tbody>
</table>
Continuous Random Variable

• A continuous random variable is one which takes an \textit{infinite} number of possible values.
• Continuous random variables are usually measurements.
• Examples:
  – height
  – weight
  – the amount of sugar in an orange
  – the time required to run a mile.
Continuous Random Variables

A continuous random variable $X$ takes on all values in an interval of numbers. The probability distribution of $X$ is described by a density curve. The probability of any event is the area under the density curve and above the values of $X$ that make up the event.
Continuous Random Variables

- A continuous random variable is not defined at specific values.
- Instead, it is defined over an interval of value; however, you can calculate the probability of a range of values.
- It is very similar to z-scores and normal distribution calculations.
Example: Young Women’s Heights

The height of young women can be defined as a continuous random variable (Y) with a probability distribution is \( N(64, 2.7) \).

A. What is the probability that a randomly chosen young woman has height between 68 and 70 inches?

\[
P(68 \leq Y \leq 70) = ???
\]
Example: Young Women’s Heights

The height of young women can be defined as a continuous random variable (Y) with a probability distribution is N(64, 2.7).

A. What is the probability that a randomly chosen young woman has height between 68 and 70 inches?

Normalcdf (68, 70, 64, 2.7) = 0.56

There is about a 5.6% chance that a randomly chosen young woman has a height between 68 and 70 inches.
Example: Young Women’s Heights

The height of young women can be defined as a continuous random variable (Y) with a probability distribution is N(64, 2.7).

B. At 70 inches tall, is Mrs. Daniel unusually tall?
Example: Young Women’s Heights

The height of young women can be defined as a continuous random variable (Y) with a probability distribution is $N(64, 2.7)$.

B. At 70 inches tall, is Mrs. Daniel unusually tall?

$\text{Normcdf}(0, 69.999999, 64, 2.7) = 0.9868$

Yes, Mrs. Daniel is unusually tall because 98.68% of the population is shorter than her.