Section 3.1 – (More About) Random Variables

Suppose an experiment is conducted. A random variable is a function that assigns values to the outcomes of the experiment.

Ex: A coin is flipped resulting in either “heads” or “tails” and the random variable $X$ is defined by

\[ X(\text{heads}) = 1 \quad X(\text{tails}) = 0 \]

We say that the random variable $X$ “indicates heads”.

Ex: A fair six sided die is tossed, let the random variable $X$ indicate the event that an even number is rolled. What is $X(2)$? $X(3)$?

\[
X(2) = 1 \quad X(\text{even}) = 1 \\
X(3) = 0 \quad X(\text{odd}) = 0
\]

Def: Any random variable whose only possible values are 0 and 1 is called a Bernoulli Random Variable.
Ex: Two gas stations are located at a certain intersection. Each one has six gas pumps. Consider the experiment in which the number of pumps in use at a particular time of day is determined for each of the stations. Define a random variable

\[ X = \text{total number of pumps in use at the two stations} \]

What are the possible values of \( X \)?

\[ X = 0, 1, 2, 3, \ldots, 12 \]

(1, 1) or (2, 0) or (0, 2)

Define a random variable

\[ Y = \text{total number of pumps not currently being used at the two stations} \]

What are the possible values of \( Y \)? How are \( X \) and \( Y \) related?

\[ Y = 0, 1, 2, 3, \ldots, 12 \]

\[ X + Y = 12 \]
Ex: Suppose pump 1 at gas station number one is observed as the next customer begins to use it. Let $T =$ the length of time the customer is at the pump. What are the possible values of $T$?

\[(0, \infty)\]

There are two types of random variables:

1. **Discrete Random Variables:** A random variable whose possible values are either finite or may be listed.

2. **Continuous Random Variables:** A random variable whose possible values consists of an interval of values or several intervals of values and for which the probability of any one value is zero.

i.e. $P(X = c) = 0$ for all $c$

For continuous random variables $X$, they take on all values in an interval of numbers. In fact, the probability of $X$ equaling an individual number is 0! The probability distribution of $X$ is described by a density curve. The probability of any event is the area under the curve and above the values of $X$ that make up the event.

Ex: Which of the above variables were discrete? Which were continuous?

$X, Y, T$
Section 3.2 - Probability Distributions for Discrete Random Variables

Def: The probability mass function (pmf) of a discrete rv is defined for every number \( x \) by

\[
p(x) = P(X = x)
\]

Ex: Six lots of components are ready to be shipped by a certain supplier. The number of defective components in each lot is as follows:

<table>
<thead>
<tr>
<th>Lot</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of defectives</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>0</td>
</tr>
</tbody>
</table>

One of these lots is to be randomly selected for shipment to a particular customer. Let \( X \) = number of defectives in the selected lot.

What are the possible values of \( X \)? Determine the values of the pmf.

\[
X = 0, 1, 2
\]

\[
\begin{array}{c|c|c|c|c|}
X & 0 & 1 & 2 \\
\hline
p(x) & \frac{3}{6} = \frac{1}{2} & \frac{1}{6} & \frac{2}{6} = \frac{1}{3} \\
\end{array}
\]

\[
\frac{1}{2} + \frac{1}{6} + \frac{1}{3} = 1 \checkmark
\]
Properties of $p$

1. $p(x) \geq 0$ for all $x \in \mathbb{R}$

2. $\sum_x p(x) = 1$ \hspace{1cm} (sum of all $p(x) = 1$)

3. $P(X \in A) = \sum_{x \in A} p(x)$, where $A \subseteq \mathbb{R}$ is a discrete set

Suppose we toss a fair coin 10 times.
Let $X =$ number of heads in the 10 tosses.
What are the possible values of $X$?

$X = 0, 1, 2, \ldots, 10$

How many heads do we expect to get in the 10 tosses? 5

prob. of heads $= \frac{1}{2}$
prob. of tails $= \frac{1}{2}$

What is $p(0)$? $p(1)$?

$p(10) = p(0) = p(X = 0) = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} \cdot \ldots \cdot \frac{1}{2} = \left(\frac{1}{2}\right)^{10} = \frac{1}{1024}$

$p(1) = p(X = 1) = 10 \left(\frac{1}{2}\right)^{10} = \frac{10}{1024}$

$\frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} \ldots \ldots \ldots \ldots \cdot \frac{1}{2}$

$\text{HTTTTTTTTTTT or THTTTTTTTTT or TTHTTTTTTTTTT}$
Parameters in probability distributions:

It is often beneficial to describe a class of probability distribution which have the same form using a single pmf by inserting a parameter. 

\[ \text{only } 0 \text{ or } 1 \]

For example, suppose you have a Bernoulli random variable, but you don’t know that probabilities associated with it. What would the pmf be?

\[ p(x) = \begin{cases} 
0 & \text{if } x = 0 \\
1 - a & \text{if } x = 0 \\
a & \text{otherwise}
\end{cases} \]

Note, this pmf describes every possible Bernoulli rv.

ex. 6 sided fair die

\[ X = 1 \text{ if roll 4} \]

\[ X = 0 \text{ if any other } \]

\[ p(x) = \begin{cases} 
\frac{1}{6} & \text{if } x = 1 \\
\frac{5}{6} & \text{if } x = 0 \\
0 & \text{otherwise}
\end{cases} \]
The Cumulative Distribution Function: 

Def: The cumulative distribution function (cdf) \( F(x) \) of a discrete rv \( X \) with pmf \( p(x) \) is defined for every number \( x \) by

\[
F(x) = P(X \leq x)
\]

For any number \( x \), \( F(x) \) is the probability that the observed value of \( X \) will be at most \( x \).

Ex: A store carries flash drives with either 1 GB, 2 GB, 4 GB, 8 GB, or 16 GB of memory. The table below gives the distribution of the rv 
\[ X = \text{the amount of memory in a purchased drive.} \]

<table>
<thead>
<tr>
<th>( X )</th>
<th>1</th>
<th>2</th>
<th>4</th>
<th>8</th>
<th>16</th>
</tr>
</thead>
<tbody>
<tr>
<td>( p(x) )</td>
<td>0.05</td>
<td>0.10</td>
<td>0.35</td>
<td>0.40</td>
<td>0.10</td>
</tr>
</tbody>
</table>

Determine the values of \( F \) for the possible values of \( X \).

\[
F(1) = P(X \leq 1) = 0.05
\]

\[
F(2) = P(X \leq 2) = 0.05 + 0.10 = 0.15
\]

\[
F(4) = P(X \leq 4) = 0.05 + 0.10 + 0.35 = 0.5
\]

\[
F(8) = P(X \leq 8) = 0.9
\]

\[
F(16) = P(X \leq 16) = 1
\]

What is the value of \( F(2.5) \)?

\[
F(2.5) = P(X \leq 2.5) = 0.15
\]

Sketch a graph of \( F \).
Example:

Suppose the random variable $X$ takes on possible values $x = 0, 1, 2, 3$ and has pmf given by $p(x) = \frac{x+1}{k}$, determine the value of $k$.

\[
\begin{array}{c|cccc}
X & 0 & 1 & 2 & 3 \\
p(x) & \frac{1}{k} & \frac{2}{k} & \frac{3}{k} & \frac{4}{k} \\
\end{array}
\]

\[
\frac{1}{k} + \frac{2}{k} + \frac{3}{k} + \frac{4}{k} = 1
\]

\[
\frac{10}{k} = 1
\]

\[
k = 10
\]
Example:

The probabilities that a customer selects 1, 2, 3, 4, or 5 items at a convenience store are 0.32, 0.12, 0.23, 0.18, and 0.15, respectively.

Construct a probability distribution for the data

<table>
<thead>
<tr>
<th>X</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>P(X)</td>
<td>0.32</td>
<td>0.12</td>
<td>0.23</td>
<td>0.18</td>
<td>0.15</td>
</tr>
</tbody>
</table>

a. Find P(X > 3.5)  
   \[ 0.18 + 0.15 = 0.33 \]

b. Find P(1.0 < X < 3.0)  
   \[ 0.12 \]

c. Find P(X < 5)  
   \[ 1 - 0.15 = 0.85 \]
Section 3.3 - Expected Values

Consider the table below which gives the number of years required to obtain a Bachelor’s degree for graduates of high school A, and the number of students who needed each:

<table>
<thead>
<tr>
<th>Years (X)</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Students</td>
<td>17</td>
<td>23</td>
<td>38</td>
<td>19</td>
<td>97</td>
</tr>
<tr>
<td>P(X)</td>
<td>17/97</td>
<td>23/97</td>
<td>38/97</td>
<td>19/97</td>
<td></td>
</tr>
</tbody>
</table>

How would compute the “average” number of years required by graduates of high school A?

\[
\frac{3(17) + 4(23) + 5(38) + 6(19)}{97} \approx 4.61
\]

\[
\sqrt{3 \left( \frac{17}{97} \right) + 4 \left( \frac{23}{97} \right) + 5 \left( \frac{38}{97} \right) + 6 \left( \frac{19}{97} \right)}
\]

The “average” value of a rv \( X \) is called the “expected value” of \( X \).

Def: Let \( X \) be a discrete rv with set of possible values \( D \) and pmf \( p \). The expected value or mean value of \( X \), denoted \( E[X] \) or \( \mu_X \) or just \( \mu \) is

\[
E[X] = \sum_{x \in D} x \cdot p(x)
\]
Properties of Expected Value and Variance

1. \( E[c] = c \) for any constant \( c \in \mathbb{R} \)

2. \( E[aX + bY] = aE[X] + bE[Y] \)

3. \( E[h(X)] = \sum_{x \in D} h(x) p(x) \)

4. \( V(X) = E[(X - \mu)^2] \) or \( V(X) = E[X^2] - E[X]^2 \)

5. \( V(aX + b) = a^2 V(X) \)

\[ E[2] = 2 \]

\[ E[X^2] = \sum x^2 p(x) \]

\[ V(X) = E[(X - \mu)^2] \]

\[ V(aX + b) = a^2 V(X) \]

\[ \text{does not impact variance} \]
Ex: Let \( X \) have pmf. given by

\[
\begin{array}{c|cccc}
 x & 1 & 2 & 3 & 4 \\
 \hline
 f(x) & 0.4 & 0.2 & 0.3 & 0.1 \\
\end{array}
\]

Determine \( E[X] \), \( E[X^2] \), and use the formula \( \sigma^2 = E[X^2] - E[X]^2 \) to determine the standard deviation of \( X \).

\[
E[X] = 1(0.4) + 2(0.2) + 3(0.3) + 4(0.1) = 2.1
\]

\[
E[X^2] = 1(0.4) + 4(0.2) + 9(0.3) + 16(0.1) = 5.5
\]

\[
\sigma_X = \sqrt{1.09} \approx 1.044
\]

Determine the expected value and variance of the rv \( Y \) defined by \( Y = 5X - 1 \), where \( X \) is given in the previous problem.

\[
E[Y] = E[5X-1] = 5E[X] - 1 = 5(2.1) - 1 = 9.5
\]

\[
\sigma_Y = \sqrt{5^2 \cdot \sigma_X^2} = \sqrt{25 \cdot 1.09} = 27.25
\]
Example:

The following distribution shows the amount of money Jane earns when baby-sitting. Let $X$ be the amount she makes for the job.

<table>
<thead>
<tr>
<th>$X$</th>
<th>20</th>
<th>25</th>
<th>30</th>
<th>35</th>
<th>40</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P(X)$</td>
<td>.15</td>
<td>.25</td>
<td>.40</td>
<td>.10</td>
<td>.10</td>
</tr>
</tbody>
</table>

$E[X^2] = \sum X^2 P(X) = 400(.15) + 625(.25) + 900(.40) + 1225(.10) + 1600(.10) = 858.75$

a. What amount can Jane “expect” to make per job?

$$E[X] = 20(.15) + 25(.25) + 30(.40) + 35(.10) + 40(.10) = 28.75$$

$$\mu[X] = 858.75 - 28.75^2 = 32.1875$$

b. What is the standard deviation Jane’s earnings?

$$\sigma_X = \sqrt{32.1875} \approx 5.6734$$

The family Jane baby-sits for decides to give her a raise. To calculate how much she will now make per job they will use the following formula:

$$\text{Amount} = 1.25(X) + 5$$

c. What amount can Jane NOW “expect” to make per job?

$$E[1.25X + 5] = 1.25 E[X] + 5 = 1.25(28.75) + 5 = 40.94$$

d. What is the standard deviation of her earnings after the raise?

$$\sigma_{1.25X+5} = 1.25 \sigma_X = 1.25(5.6734) \approx 7.092$$