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

$$X(\text{heads}) = 1 \qquad X(\text{tails}) = 0$$

We says 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 $X(2)$? $X(3)$?

$X(2) = 1$ $\qquad$ $X(\text{even}) = 1$

$X(3) = 0$ $\qquad$ $X(\text{odd}) = 0$

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 = 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 = 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?

$\underbrace{(0, \infty)}_{\text{interval}}$

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:

Lot	1	2	3	4	5	6
Number of defectives	0	2	0	1	2	0

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.

$X = 0, 1, 2$

x	0	1	2
$p(x)$	$3/6 = 1/2$	$1/6$	$2/6 = 1/3$

$1/2 + 1/6 + 1/3 = 1$ ✓

Properties of p

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

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

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

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

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

same as 10 tails

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

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

H T T T T T T T T T or T H T T T T T T T T or T T H T T T T T T T 10

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

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*.

→ only 0 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?

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

Note, this pmf describes *every* possible Bernoulli rv.

ex: 6 sided fair die

$X = 1$ if roll 4

$X = 0$ if any other #

$$p(x) = \begin{cases} 1/6 & \text{if } X = 1 \\ 5/6 & \text{if } X = 0 \\ 0 & \text{otherwise} \end{cases}$$

The Cummulative Distribution Function: pmf $p(X=x)$

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

$$F(x) = P(X \le x)$$

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 = the amount of memory in a purchased drive.

pmf :

X	1	2	4	8	16
$p(x)$	0.05	0.10	0.35	0.40	0.10

Determine the values of F for the possible values of X.

$F(1) = P(X \le 1) = .05$

$F(2) = P(X \le 2) = .05 + .10 = .15$

$F(4) = P(X \le 4) = .05 + .10 + .35 = .5$

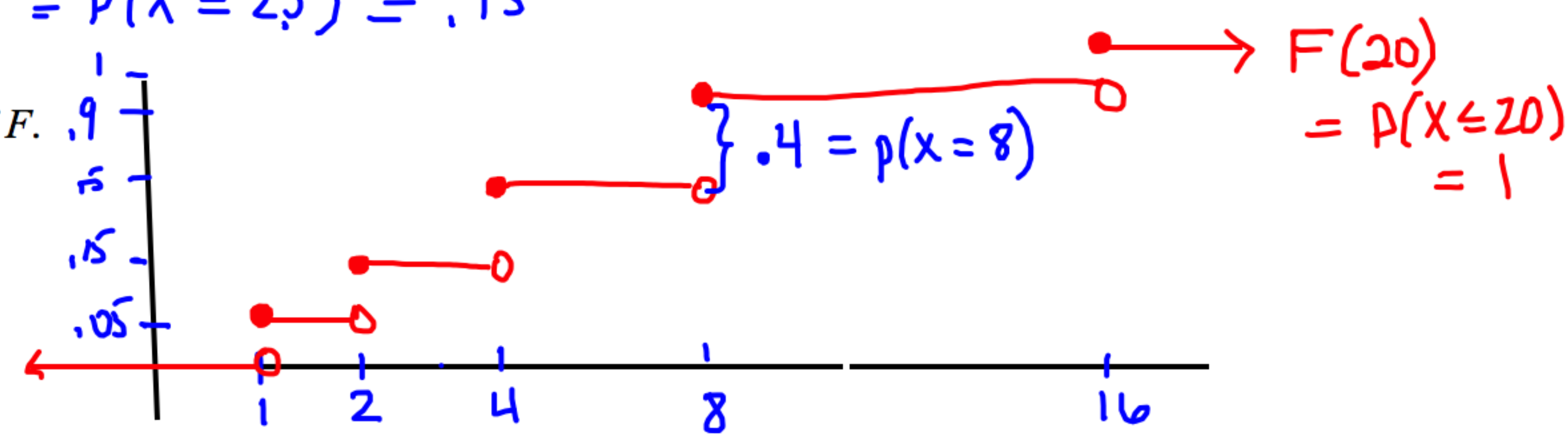
$F(8) = P(X \le 8) = .9$

$F(16) = P(X \le 16) = 1$

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

$F(2.5) = P(X \le 2.5) = .15$

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.

X	0	1	2	3
p(x)	$\frac{1}{K}$	$\frac{2}{K}$	$\frac{3}{K}$	$\frac{4}{K}$

$$\frac{1}{K} + \frac{2}{K} + \frac{3}{K} + \frac{4}{K} = 1$$

$$\frac{10}{K} = 1$$

$$\boxed{K = 10}$$

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

X	1	2	3	4	5
P(X)	.32	.12	.23	.18	.15

a. Find P(X > 3.5)

.18 + .15 = .33

b. Find P(1.0 < X < 3.0)

.12

c. Find P(X < 5)

1 − .15 = .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:

Years (X)	3	4	5	6	total
Number of Students	17	23	38	19	97
p(X)	17/97	23/97	38/97	19/97	

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

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

$$3\left(\tfrac{17}{97}\right)+4\left(\tfrac{23}{97}\right)+5\left(\tfrac{38}{97}\right)+6\left(\tfrac{19}{97}\right)$$

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 μ_X or just μ is $E[X]=\underbrace{\sum_{x\in D}}_{\text{Sum}} x\cdot p(x)$

Properties of Expected Value and Variance

1. $E[c] = c$ for any constant $c \in \mathbb{R}$ $\quad E[2] = 2$

2. $E[aX + bY] = aE[X] + bE[Y]$

3. $E[h(X)] = \sum_{x \in D} \underline{h(x)}\, p(x)$ $\quad E[X^2] = \sum X^2 p(X)$

4. $V(X) = E\left[\underline{(X-\mu)^2}\right]$ or $V(X) = E[X^2] - E[X]^2$

5. $V(aX + \underline{b}) = a^2 V(X)$

$\uparrow$ does not impact variance

Ex: Let X have pmf. given by

x^2	1	4	9	16
x	1	2	3	4
$f(x)$	0.4	0.2	0.3	0.1

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.

$$E[X] = 1(.4) + 2(.2) + 3(.3) + 4(.1) = 2.1$$

$$E[X^2] = 1(.4) + 4(.2) + 9(.3) + 16(.1) = 5.5$$

$$V[X] = \sigma_X^2 = 5.5 - (2.1)^2 = 1.09 \qquad \sigma_X = \sqrt{1.09} \approx 1.044$$

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

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

$$V[Y] = V[5X-1] = 5^2 V[X] = 25(1.09) = 27.25$$

Example:

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

X^2	400	625	900	1225	1600
X	20	25	30	35	40
P(X)	.15	.25	.40	.10	.10

$E[X^2] = 858.75$

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

$E[X] = 20(.15) + 25(.25) + 30(.40) + 35(.10) + 4(.10) = \$28.75$

$V[X] = 858.75 - 28.75^2 = 32.1875$

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:

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

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

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

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

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