Chapter 7
Random Variables

In Chapter 6, we learned that a "random phenomenon" was one that was unpredictable in the short term, but displayed a predictable pattern in the long run. In Statistics, we are often interested in numerical outcomes of random phenomena. In this chapter, we will learn to define random variables to describe numeric outcomes of random phenomena as well as how to calculate the means and variances of such random variables.

Objectives:

- Define what is meant by a random variable
- Define a discrete random variable
- Define a continuous random variable
- Explain what is meant by the probability distribution for a random variable
- Explain what is meant by the law of large numbers
- Calculate the mean and variance of a discrete random variable
- Calculate the mean and variance of distributions formed by combining two random variables

Case Study:

Jane Blaylock joined the LPGA in 1969 and by 1972had become the leading money winner on the tour. During the Bluegrass Invitational Tournament she was disqualified for an alleged rules infraction. The LPGA appointed a committee of her competitors, who suspended her from her next tournament, the Carling Open. Blaylock sued for damages and expenses under the Sherman Antitrust Act, which says that individuals cannot be prevented by their peers from working in their profession because it would lessen competition. She won but then had to come up with a method of determining how much money she might reasonably have made if she had played the Open. This is a difficult issue for most antitrust cases but was particularly problematic for a professional golfer who might play well one day and poorly another day. Her task was challenging. She had to use a measure that the judge and jury would understand and that would be convincing for a ruling in her favor. She and her legal team used a statistical procedure called the expected value. Using data from the nine most recent tournaments that she had played in prior to the disqualification, they estimated the probability that she would achieve various scores based on her past performance. The scores for the players who won money ranged from 209 (the tournament winner) to 232. To simplify things, the 24 possible scores were reduced to 8, where 210, for instance, represents 209, 210, and 211. Here is a table that summarizes her possible outcomes and the probabilities calculated for each:

Possible outcomes X	210	213	216	219	222	225	228	231
Probability, P(X)	0.07	0.16	0.23	0.24	0.17	0.09	0.03	0.01

The probability is 0.07 that she would score 209, 210, or 211 and so forth. Using these numbers her expected score was calculated to be approximately 218. Had she played in the tournament, her 218 would have earned her $1427.50. Not only was the jury persuaded, but they also believed that Blaylock might well have won the tournament, so they awarded her first-place money, \$4500. This amount was then tripled to \$13,500 to cover legal expenses, according to the provisions of the Sherman act. Statistics to the rescue!

When we roll two dice, we can define the sum to be a random variable; *X*. *X* can take on any value from 2 through 12. Since we don't know exactly what sum will appear on a given roll, we call *X* a *random variable.*

Random Variable:
A random variable is a variable whose value is a numerical outcome of a random phenomena

Example 1: Put all the letters of the alphabet in a hat. If you choose a consonant, I pay you \$1. If you choose a vowel, I pay you \$5. *X* is the random variable representing the outcome of the experiment. What are the possible values of X?

$x = 1$ or 5

Example 2: In most college courses, you get as a grade, either an A, B, C, D, or F. For credit purposes, A's are given 4 points, B's are given 3 points, C's are given 2 points, Ds are given 1 point, and F's are no points. Let *X* be the random variable representing the points a student gets. What are the possible values of X?

$X = 0, 1, 2, 3, 4$

Example 3: You are tossing 5 coins. Let X be the random variable representing the number of heads. What are the possible values of X?

$X = 0, 1, 2, 3, 4, 5$

Discrete Random Variable
A discrete random variable has a countable number of possible values

Probability Distribution:

lists the values of the random variable X and their probabilities

$X=$	x_1	x_2	x_3	...	x_n
$P(X)$	P_1	P_2	P_3	...	p_n

The probabilities P_x must: 1) $0 \le P_n \le 1$ 2) $p_1 + p_2 + p_3 + \dots p_n = 1$

Example 4: A college course has the following grade distributions.

Grade, X	0	1	2	3	4
P(X)	0.01	0.05	0.30	0.43	0.21

Verify that it is a legitimate probability distribution.

All probabilities are between zero and 1

$.01 + .05 + .30 + .43 + .21 = 1$

Example 5: Put all the letters of the alphabet in a hat. If you choose a consonant, I pay you $1. If you choose a vowel, I pay you $5. *X* is the random variable representing the outcome of the experiment. Create the distribution of *X*.

X = Money won	$1	$5
P(X)	21/26	5/26

Example 6: A college instructor teaching a large class traditionally gives 10% A's, 20% B's, 45% C's, 15% D's, and 10% F's. If a student is chosen at random from the class, the student's grade on a 4-point scale (A = 4) is a random variable *X*. Create the distribution of *X*.

X = Grade pts	0	1	2	3	4
P(X)	.1	.15	.45	.2	.1

What is the probability that a student has a grade point of 3 or better in this class?

$P(X \ge 3) = .3$

What is the probability that a student has a grade point of 2 or worse in this class?

$P(X \le 2) = .7$

Example 7: Consider rolling two dice. Define X to be the sum of the two dice. Construct the probability distribution of X and display it with a **probability histogram**.

Example 8: Consider flipping a coin 4 times and recording H or T. Define X to be the number of Heads flipped. Construct the probability distribution of X and use it to answer the following questions:

P(X=0)= $\frac{1}{16}$

P(X=0 or X = 1)= $5/16$

P(X>2)= $5/16$

P(X ≤ 3)= $15/16$

P(X ≥ 1)= $15/16$

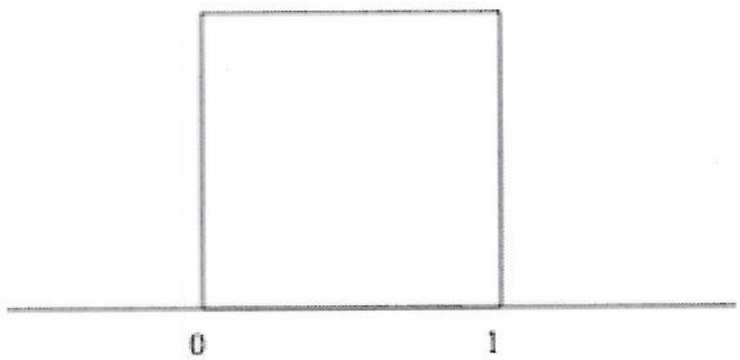
X: # of Heads	0	1	2	3	4
P(x)	$1/16$	$\frac{4}{16}$	$\frac{6}{16}$	$\frac{4}{16}$	$1/16$

Continuous Random Variables

Rolling dice and flipping coins result in random variables whose outcomes are *countable*. Some situations result in outcomes that can take on any value over a given interval.

Continuous Random Variable:

Example 8: A random number is chosen from 0 to 1. Answer the following questions:



a. Find $P(X > .5)$

b. Find $P(X \geq .5)$

c. Find $P(X \leq .8)$

d. Find $P(X \geq .15)$

e. Find $P(.2 < X < .4)$

f. Find $P(.2 < X \leq .4)$

g. Find $P(X < 1)$

h. Find $P(X = .5)$

Note: Continuous Models assign probabilities to intervals, NOT individual values of X! The $P(X = 2) = 0$. Also, $P(X > 3) = P(X \geq 3)$

Example 9: Suppose a random variable is uniformly distributed between the values $3<X<7$. Sketch the density curve describing this variable's probability distribution.

a) Find the height of the uniform distribution

b) Find $P(X>4.5)$

c) Find $P(3.2<X<6.4)$

d) Find P(X"5.2)

Example 10: Choose an American household at random and let the random variable X be the number of persons living in the household. The probability of X is:

X	1	2	3	4	5	6	7 or more
P(X)	0.25	0.32	0.17	0.15	0.07	0.03	0.01

a. Find $P(X > 4)$

.11

b. Find $P(X \geq 4)$

.26

c. Find $P(2 < X \leq 4)$

.32

d. Find $P(X \neq 2)$

.68

Example 11: A fair coin is flipped 4 times. Find the probability distribution of the discrete random variable X that counts the number of heads.

X = # of Heads	0	1	2	3	4
P(x)	$\frac{1}{16}$	$\frac{4}{16}$	$\frac{6}{16}$	$\frac{4}{16}$	$\frac{1}{16}$

a. Find $P(X > 2)$

$\frac{5}{16}$

b. Find $P(X \geq 2)$

$\frac{11}{16}$

c. Find $P(X \geq 1)$

$\frac{15}{16}$

d. Find $P(X \geq 0)$

1

Example 12: Suppose each of four randomly selected customers purchasing a monkey shaving kit at a certain store chooses either an electric (E) or a gas (G) model. Assume that these customers make their choices independently of one another and that, according to Consumer Reports, 40% of all customers select an electric model while 60% tend to choose the gas model. One possible purchase outcome for the four is EGGE. Mmmmmmm...."egge". Find the probability that at least two of the four customers will choose the electric model. Be sure to define your random variable, provide a probability distribution and answer the question.

Example 13: A coin that is rigged is flipped 3 times. The probability of heads is twice the probability of tails. Find the probability distribution of the discrete random variable X that counts the number of heads.

$P(H) = \frac{2}{3}$ $P(T) = \frac{1}{3}$

X = # of Heads	0	1	2	3
$P(x)$	$\frac{1}{27}$	$\frac{6}{27}$	$\frac{12}{27}$	$\frac{8}{27}$

$3 \quad \frac{2}{3} \cdot \left(\frac{1}{3}\right)^2 \quad 3 \left(\frac{2}{3}\right)^2 \cdot \frac{1}{3}$

a. Find $P(X > 2)$ $8/27$

b. Find $P(X \geq 2)$ $20/27$

c. Find $P(X \leq 3)$ 1

d. Find $P(X \neq 2)$ $15/27$

Example 14: The numbers 1, 2, 3, 4, 5 are placed in a hat. A number is chosen at random, replaced, and another number is chosen at random. Let X be the number of odd numbers that are chosen. Find the probability distribution of the random variable X.

$P(\text{odd}) = 3/5$

$P(\text{even}) = 2/5$

X = # of odd	0	1	2
$P(x)$	$\frac{4}{25}$	$\frac{12}{25}$	$\frac{9}{25}$

a. Find $P(X > 1)$ $\frac{9}{25}$

b. Find $P(X \geq 1)$ $\frac{21}{25}$

c. Find $P(X \leq 1)$ $\frac{16}{25}$

d. Find $P(X \neq 2)$ $16/25$

Example 15: An SRS of 3 students are chosen to be on a committee that advises a university on its computer policies. 70% of the student population is PC users and 30% are Mac users. Let X be the number of PC users on the student committee. Find the probability distribution of the random variable X.

X = # of PC	0	1	2	3
$P(x)$	$.3^3 =$ $.027$	$3(.3)^2(.7)$ $.189$	$3(.3)(.7)^2$ $.441$	$.7^3$ $.343$

a. Find $P(X > 2)$

$.343$

b. Find $P(X \geq 2)$

$.784$

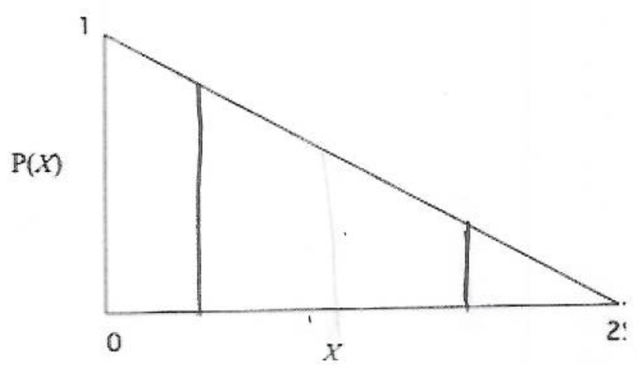
c. Find $P(X \leq 1)$

$.216$

d. Find $P(X \neq 3)$

$.657$

Example 16. A density curve is shown. Find the following probabilities.



a. Find $P(X > 1)$

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

$\frac{1}{4}$

b. Find $P(X < 1.5)$

$1 - \frac{1}{2}\left(\frac{1}{2}\right)\left(\frac{1}{4}\right)$

$1 - \frac{1}{16}$

$\frac{15}{16}$

c. Find $P(.5 \leq X \leq 1.5)$

$\frac{1}{2}\left(\frac{3}{2}\right)\left(\frac{3}{4}\right) - \frac{1}{16}$

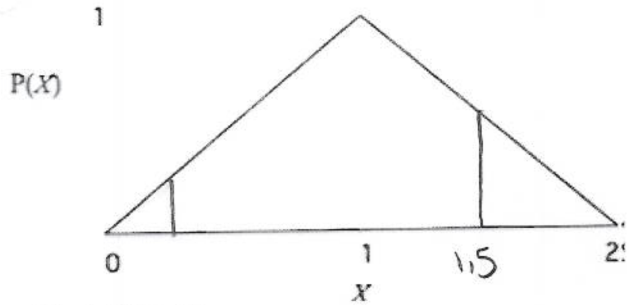
$\frac{9}{16} - \frac{1}{16}$

$\frac{1}{2}$

d. Find $P(X < 2)$

1

Example 17: Two random numbers from 0 to 1 are chosen and added. Let X be their sum. X is thus a continuous random variable between 0 and 2. The density curve is shown. Find the following probabilities:



a. Find $P(X > 1)$

$.5$

b. Find $P(X > 1.5)$

$\frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8}$

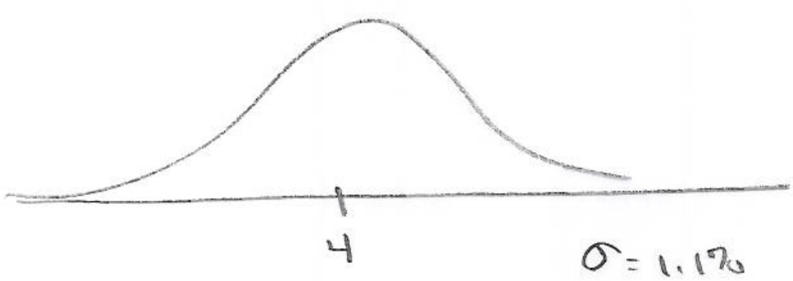
c. Find $P(X < .2)$

$\frac{1}{2}(.2)(.2) = .02$

d. Find $P(.5 \le X \le 1.5)$

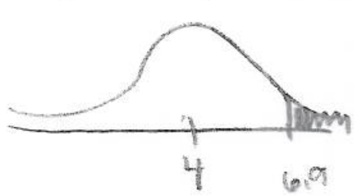
$1 - \frac{1}{8} - \frac{1}{8} = 3/4$

Example 18: Accurate labeling of packaged meat is difficult because of weight decrease due to moisture loss (defined as a percentage of the package's original net weight). Suppose that moisture loss for a package of chicken breasts is N(4%, 1.1%) as suggested in the paper "Drained Weight Labeling for Meat and Poultry: An Economic Analysis of a Regulatory Proposal" (J. of Consumer Affairs (1980): 307-325). Let X denote the moisture loss for a randomly selected package.



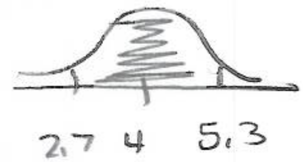
a)What is the probability X is at most 4.2%? $P(x \le 4.2) = .572$

b)What is the probability X is at least 6.9%? $P(x \ge 6.9) = .0042$



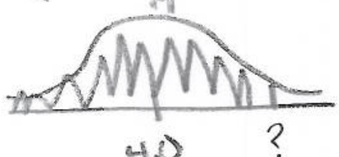
c) What is the probability moisture loss differs from the mean value by at least 1.3%

$P(2.7 \le X \le 5.3) = .7627$



d) Find a moisture loss value such that 90% of packages have a loss below that value.

$P(X \le \quad) = .9$



5.41%

Example 19: According to a recent AP poll, approximately 40% of American adults indicated they used the internet to get news and information about political candidates. Suppose 40% of all American adults use this method to get their political information. What would happen if you randomly sampled a group of 1500 American adults and asked them if they used the internet to get this information? Define X to be the % of your sample that would respond that the internet was their primary source. We will learn in chapter 9 that the distribution of X is approximately N(0.4, 0.01265). Use this information to sketch the probability distribution of X and answer the following questions:

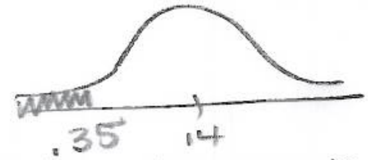
If you conducted a survey of 1500 American Adults, what % would you expect use the internet as their primary source?

What is $P(X \ge 0.42)$?

$\sigma = .01265$

.0569

What is $P(X \le 0.35)$? .080039



What is P(your result is within 5% of the *actual* % who use the internet as a primary source)?

$P(.35 \le X \le .45) = \quad .9999$