

AP Stats Chapter 7 Notes: Random Variables

If we were to toss a coin we could have two outcomes (heads or tails). If we tossed a coin four times we might end up with a sample space of HTTH. If we let X be the number of heads, it would be 2. If our outcome was HTTT then $X=1$. The possible values of X are 0,1,2,3, and 4. Tossing a coin four times will give us one value of X , doing so again will give us another value of X . We call X a **random variable** because its values vary when the coin tossing is repeated.

Random Variable:

Discrete and Continuous Random Variables

There were several rules of probability in Chapter 6, but the basic understanding is that the outcome probabilities must be between 0 and 1 and have sum 1. When the outcomes are numerical, then they are values of a random variable. When random variables have probabilities assigned they are called **discrete random variables**.

Discrete Random Variable:

Example 7.1 Getting Good Grades

North Carolina State University posts the grade distribution for its courses online. Students in Statistics 101 in the fall 2003 semester received 21% A's, 43% B's, 30% C's, 5% D's and 1% F's. Choose a Statistics 101 student at random. To choose at random means give every student the same chance to be chosen. The student's grade on a four-point scale (with A=4) is a random variable.

The value of X changes when we repeatedly choose students at random, but it is always one of 0, 1, 2, 3, or 4. Here is the distribution of X .

Value of X :	0	1	2	3	4
Probability:	0.01	0.05	0.30	0.43	0.21

The probability that a student got a B or better is the sum of the probabilities of an A and a B. Written differently:

Probability Histograms

Example 7.2 Tossing Coins

What is the probability distribution of the discrete random variable X that counts the number of heads in four tosses of a coin? We can derive this distribution if we make two reasonable assumptions:

1. The coin is balanced, so each toss is equally likely to give H or T.
2. The coin has no memory, so tosses are independent.

The outcome of four tosses in a sequence of heads and tails such as HTTH. There are 16 possible outcomes in all. What are the possible outcomes?

Each of these outcomes has an equal chance of occurring (1/16).

Let X be the number of heads. What values may this discrete random variable take on?

Are these values equally likely?

$P(X=0) =$

$P(X=1) =$

$P(X=2) =$

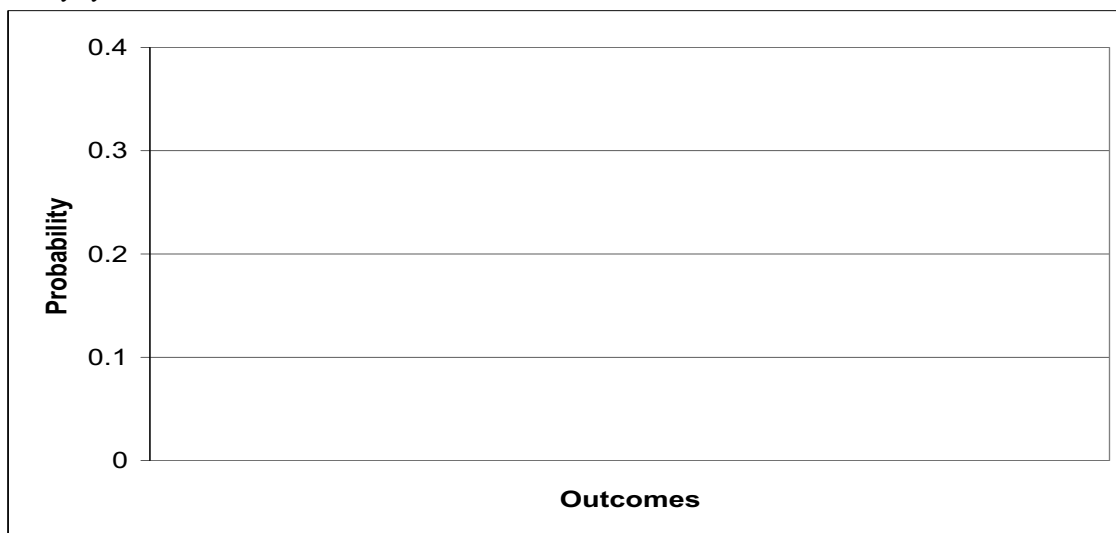
$P(X=3) =$

$P(X=4) =$

In table format:

Number of heads	0	1	2	3	4
Probability					

Draw a probability histogram for this distribution. The probability distribution is exactly symmetric. It is an idealization of the relative frequency distribution of the number of heads after many tosses of four coins, which would be nearly symmetric but is unlikely to be exactly symmetric.



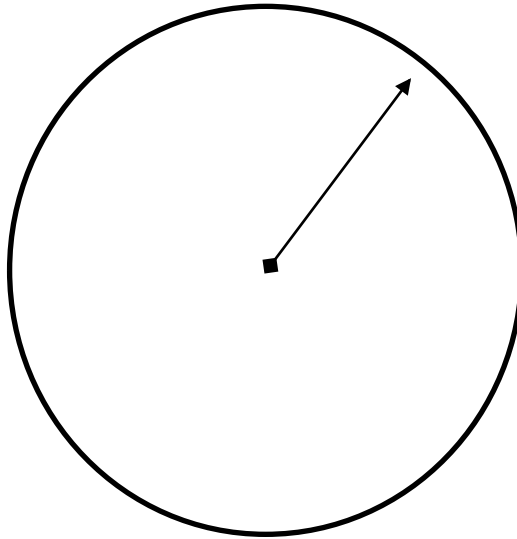
Probability of at most two heads?

Probability of at least one head?

Assignment: p. 469-470 7.1 to 7.6

Continuous Random Variables

When we use the table of random numbers to select a digit between 0 and 9, the result is a discrete random variable. The probability model assigns probability $1/10$ to each of the 10 possible outcomes (0 to 9). Suppose that we want to choose a number at random between 0 and 1, allowing any number between 0 and 1 as the outcome. Think about a spinner:



The sample space is now an entire interval of numbers. $S = \{ \hspace{10em} \}$

How can we assign probabilities to such events as 0.3 to 0.7? With random digits we know all outcomes are equally likely, and in this case we would like all possible outcomes to be equally likely. But we can't assign probabilities to each individual value of x and then sum, because there are infinitely many possible values. Instead, we use areas under a density curve. Density curves have an area of 1 underneath, corresponding to total probability of 1.

Density Curve Example:

X in the above example would be a continuous random variable, not a discrete random variable. This is true since the values of X are an interval of numbers and not specific numbers.

Continuous Random Variable: X takes 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.

The probability model for a continuous random variable assigns probabilities to intervals of outcomes rather than to individual outcomes. In fact, **all continuous probability distributions assign probability 0 to every individual outcome**. Only intervals of values have positive probability.

Normal Distributions as Probability Distributions

Because any density curve describes an assignment of probabilities, Normal Distributions are probability distributions. Recall $N(\mu, \sigma)$ is shorthand for a Normal distribution with mean μ and standard deviation σ . If X has the $N(\mu, \sigma)$ distribution, then

$Z = \frac{X - \mu}{\sigma}$ is a standard Normal random variable having the distribution N (0, 1).

Example 7.4

A sample of 400 college students was asked a question about cheating. 12% said if they witnessed cheating, they would report it to the professor. Suppose that if we could ask all college students, 12% would answer Yes.

The proportion $p = 0.12$ is a *parameter* that describes the population of all college students. The proportion p of the sample who answer Yes is a *statistic* used to estimate p . The statistic p is a random variable because if the sample were repeated there would be a different 400 students and thus a different value of p .

Suppose p is approximately a normal distribution where $N(0.12, 0.016)$.

What is the probability that the survey result differs from the truth about the population by more than 2 percentage points? Because $p = 0.12$, the survey misses by 2 percentage points if $p < 0.10$ or $p > 0.14$. Standardize and then use Table A to find the area under the curve.

From Table A

Means and Variances of Random Variables

To see how means of random variables work, consider a random variable that takes values $\{1, 1, 2, 3, 5, 8\}$. Do the following:

1. Calculate the mean μ of the population.
2. Make a list of all of the samples of size 2 from this population. Notice that there are two 1's. To distinguish them from each other, one will be 1_a and the other will be 1_b . There should be 15 subsets of size 2 in your list. As you list the sets, find the mean of each set:

Sample Number	Sample	\bar{x}
1	$1_a, 1_b$	1
2	$1_a, 2$	1.5
3		
4		
5		
6		
7		
8		
9		
10		
11		
12		
13		
14		
15		

3. Now find the mean of the 15 \bar{x} values in the third column. Compare this with the population mean that you calculated in 1.

The Mean of a Random Variable

Example: Most states and Canadian provinces have government sponsored lotteries. Here is a simple lottery wager, from the Tri-State Pick 3 game that New Hampshire shares with Maine and Vermont. You choose a three digit number; the state chooses a three digit winning number at random and pays you \$500 if your number is chosen. Because there are 1000 three digit numbers, you have probability $1/1000$ of winning. Taking X to be the amount your ticket pays you, the probability distribution of X is

Payoff X :	\$0	\$500
Probability:	0.999	0.001

What is your average payoff from many tickets? If you just average the 0 and 500, the average would be \$250, but that makes no sense since \$0 is much more likely than \$500. Over time, you will only get \$500 1 once in every 1000 tickets and \$0 999 times out of 1000 tickets. So...

$$\$500 \frac{1}{1000} + \$0 \frac{999}{1000} = \$0.50$$

That number is the mean of the random variable X . You spend \$1, so over time the state keeps half of the money you wager.

Mean of a probability distribution:

Expected value/Mean of a discrete random variable:

Suppose that X is a discrete random variable whose distribution is

Value of X :	x_1	x_2	x_3	...	x_k
Probability:	p_1	p_2	p_3	...	p_k

To find the mean of X , multiply each possible value by its probability, then add all the products:

Variance of a Random Variable

Suppose the X is a discrete random variable whose distribution is

Value of X:	x_1	x_2	x_3	...	x_k
Probability:	p_1	p_2	p_3	...	p_k

and that μ is the mean of X. The **variance** of X is

The **standard deviation** σ_X of X is the square root of the variance.

Example: Linda is a sales associate at a large auto dealership. She motivates herself by using probability estimates of her sales. For a sunny Saturday in April, she estimates her car sales as follows:

Cars Sold:	0	1	2	3
Probability:	0.3	0.4	0.2	0.1

We can find the mean and variance of X by arranging the calculation in the form of a table. Both μ_x and σ_x^2 are sums of columns in the table:

x_i	p_i	$x_i p_i$	$(x_i - \mu_x)^2 p_i$
0	0.3	0.0	$(0 - 1.1)^2(0.3) = 0.363$
1			
2			
3			

$$\mu_x = 1.1$$

$$\sigma_x^2 =$$

What would the standard deviation of X be?

Assignment: p. 486-487 7.24 to 7.28

Law of Large Numbers: Draw independent observations at random from any population with finite mean μ . Decide how accurately you would like to estimate μ . As the number of observations drawn increases, the mean \bar{x} of the observed values eventually approaches the mean μ of the population as closely as you specified and then stays that close.

Insurance companies, casinos, and even fast food restaurants use the law of large numbers.

The “law of small numbers”

Write down a sequence of heads and tails that you think imitates 10 tosses of a balanced coin.

How long is your longest “run” of heads or tails?

What is the actual probability of a run of three or more heads or tails?

“Hot hands” in basketball or gambling don’t actually exist. Over time the only truth that holds is the law of large numbers. The law of small numbers is actually a myth. Your brain cannot easily distinguish what is actually random and what is not random.

Assignment: p. 490-491 7.31 to 7.34

Rules for Means

X is a variable and a and b are fixed numbers, Then

1)

- If you add a number, say 5, to each value of x then the mean is also 5 more.
- If you multiply each Value of x by a number, say 10, then the mean is also 10 times more.

Example: Changing Temperature from Fahrenheit to Celsius

2)

- If you are adding two means, the combined mean is the sum.

Example: Mean number of dimple on refrigerator doors is .7 and the mean number of sag marks is 1.4. Then the mean of combined flaws, dimples and sags $.7+1.4 = 2.1$.

Rules for Variances

If X is a random variable and a and b are fixed numbers

If X and Y are independent random variables, then

Example: The Payoff X of a \$1 ticket in the Tri-State Pick 3 game is \$500 with probability 1/1000 and \$0 the rest of the time.

Find the mean, Variance, and Standard Deviation.

Suppose you play two days in a row. Find the mean, variance and standard deviation.

Note: Variances add, standard deviations do not add.

When you buy a ticket, your winnings are $W = X - 1$ because the dollar you paid has to be subtracted from the winnings. What would the mean amount of your win be?

Under this condition, the variance and standard deviation won't change, just the mean. Subtracting a fixed number changes the mean, but not the variance.

Combining Normal Random Variables

Tom and George are playing in a golf tournament. Tom's score X has the $N(110, 10)$ distribution, and George's score Y has the $N(100, 8)$ distribution. If they play independently, what is the probability that Tom will score lower than George? The difference $X - Y$ between their scores is Normally distributed, with mean and variance: