Random Variables:
A ____________________________________________ of some chance process.

Ex: (See Introduction on page 340) In the ‘Bottled Water vs. Tap Water’ activity, 13 out of 21 students made correct identifications. If we assume that students can’t tell tap water from bottled water, then each would have a 1/3 chance of being correct just by guessing.

X = number of students who guessed correctly = 13 (__________________________) 

Probability Distribution of a random variable gives its ______________________________ and their _____________________.

Probability Distribution = Probability Model for the random variable

Facts about random variables:

- Random variables are denoted by ________________________, usually X.

- Each random variable has a __________________________ that gives information about the likelihood that a specific event happens and about what’s expected to happen if the chance behavior is repeated many times.

- Two types of random variables: ________________________________.
Random Variables

X takes a fixed set of values with gaps in between.
The probability distribution of a discrete random variable X lists the values ____ and their probabilities ____ (always ________________)

Mean (Expected Value)

Variance:

Standard Deviation

X takes all the values in an ______________ of numbers.
The probability distribution of X is described by a _______________.
The probability of any event is the area under the density curve and above the values of X that make up the event.

All continuous probability models assign probability ____ to every ______________ outcome.
Example: Apgar Scores: Babies’ Health at Birth (Discrete Random Variables)

In 1952, Dr. Virginia Apgar suggested five criteria for measuring a baby’s health at birth: skin color, heart rate, muscle tone, breathing, and response when stimulated. She developed a 0-1-2 scale to rate a new born on each of the five criteria. A baby’s Apgar score is the sum of the rating on each of the five scales which gives a whole-number value from 0 to 10. Apgar scores are still used today to evaluate the health of newborns.

What Apgar scores are typical? To find out, researchers recorded the Apgar scores of over 2 million newborn babies in a single year. Imagine selecting one of these newborns at random. Define the random variable $X = $ Apgar score of randomly selected baby one minute after birth. The table below gives the probability distribution for $X$.

<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) Show that the probability distribution for $X$ is legitimate.

b) Make a histogram of the probability distribution. Describe what you see.

c) Doctors decided that Apgar scores of 7 or higher indicate a healthy baby. What is the probability that a randomly selected baby is healthy?

d) What is the probability that a randomly selected baby has an Apgar score of more than 7?
See Ex. Winning (and Losing) at Roulette (pg.344)

Apgar Scores: What’s Typical?

<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>

e) Compute the mean of the random variable X and interpret this value in context.

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

f) Compute and interpret the standard deviation of the random variable X.

The standard deviation of X is ___________________________. On average, a randomly selected baby’s Apgar score will differ from the mean (8.128) by about _______ units.

CYU: Pg. 349
Density Curves:

Example: Young Women’s Heights (Normal Probability Distributions)

The heights of young women closely follow the Normal distribution with mean $\mu = 64$ inches and standard deviation $\sigma = 2.7$ inches. This is a distribution for a large set of data. Now choose one young woman at random. Call her height $Y$. If we repeat the random choice very many times, the distribution of values of $Y$ is the same Normal distribution that describes the heights of all young women. Find the probability that the chosen woman is between 68 and 70 inches tall.

Always start by defining the random variable of interest: $Y = \ldots$

and stating the probability you are trying to find:

$P(\ldots \leq Y \leq \ldots)$
The effects of Transformations on the shape, center, and spread of a distribution of data.

<table>
<thead>
<tr>
<th>Effects on random variables of:</th>
<th>Center and Location (mean, median, quartiles (Q_1) and (Q_3), and percentiles)</th>
<th>Spread (range, IQR, variance, and standard deviation)</th>
<th>Shape of distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adding or Subtracting a Constant (a) to each value of the random variable</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Multiplying or Dividing each value of a random variable by a Constant (b)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

CYU: Pg.362

Example: The Baby and the Bathwater (pg. 363)

One brand of bathtub comes with a dial to set the water temperature. When the ‘babysafe’ setting is selected and the tub is filled, the temperature \(X\) of the water follows a Normal distribution with a mean of 34°C and a standard deviation of 2°C.

a) Define the random variable \(Y\) to be the water temperature in degrees Fahrenheit and find the mean and standard deviation of \(Y\)

b) According to Babies R Us, the temperature of a baby’s bathwater should be between 90°F and 100°F. Find the probability that the water temperature on a randomly selected day when the ‘babysafe’ setting is used meets the recommendation. Show your work.
### Mean of the Sum of Random Variables

For any two random variables \( X \) and \( Y \), if \( T = X + Y \), then the expected value of \( T \) is:

\[ \text{Mean of the sum of several random variables is the sum of their means} \]

### Mean of the Difference of Random Variables

For any two random variables \( X \) and \( Y \), if \( D = X - Y \), then the expected value of \( D \) is:

\[ \text{Mean of the difference of several random variables is the difference of their means} \]

**The order of subtraction is important**

### Independent Random Variables

Knowing whether any event involving \( X \) alone has occurred tells us nothing about the occurrence of any event involving \( Y \) alone and vice versa.

### Variance of the Sum of Independent Random Variables

For any two independent random variable \( X \) and \( T \), if \( T = X + T \), then the variance of \( T \) is:

\[ \text{Variance of the sum of several independent random variables is the sum of their variances} \]

### Variance of the Difference of Independent Random Variables

For any two independent random variable \( X \) and \( Y \), if \( D = X - Y \), then the variance of \( D \) is:

\[ \text{Variance of the difference of several independent random variables is the sum of their variances} \]
Remember
You can add variances only if the two random variables are independent, and you can never add standard deviations.

Example: Give Me Some Sugar!

Mr. Starnes likes sugar in his hot tea. From experience, he needs between 8.5 and 9 grams of sugar in a cup of tea for the drink to taste right. While making his tea one morning, Mr. Starnes adds four randomly selected packets of sugar. Suppose the amount of sugar in these packets follows a Normal distribution with mean 2.17 grams and standard deviation 0.08 grams.

What is the probability that Mr. Starnes’s tea tastes right?
Two special types of random variables:

<table>
<thead>
<tr>
<th>Setting</th>
<th>Setting</th>
</tr>
</thead>
</table>
| Arises when we perform several independent trials of the same chance process and record the outcome that a particular outcome occurs. The four conditions for a binomial setting are:  
  B: ______________? The possible outcomes of each trial can be classified as “success” or “failure”  
  I: ______________? Trials must be independent; that is, knowing the result of one trial must not have any effect on the result of any other trial.  
  N: ______________? The number of trials ‘\( n \)’ of the chance process must be fixed in advanced.  
  S: ______________? On each trial, the probability \( p \) of success must be the same.  |
| Arises when we perform independent trials of the same chance process and record the number of trials until a particular outcome occurs. The four conditions for a geometric setting are:  
  B: ______________? The possible outcomes of each trial can be classified as “success” and “failure”  
  I: ______________? Trials must be independent; that is, knowing the result of one trial must not have any effect on the result of any other trial.  
  T: ______________? The goal is to count the number of trials until the first success occurs.  
  S: ______________? On each trial, the probability \( p \) of success must be the same.  |

Binomial Random Variable: The count \( X \) of successes in a binomial setting.

The probability distribution of \( X \) is a binomial distribution with parameters \( n \) and \( p \), where \( n \) is the number of trials of the chance process and \( p \) is the probability of a success on any one trial. The possible values of \( X \) are the whole numbers from 0 to \( n \).  

Geometric Random Variable: The number of trials \( Y \) that it takes to get a success in a geometric setting.

The probability distribution of \( Y \) is a geometric distribution with parameter \( p \), the probability of a success on any trial. The possible values of \( Y \) are 1, 2, 3, …
Binomial Coefficient:
The number of ways of arranging k successes among n observations is given by the binomial coefficient

<table>
<thead>
<tr>
<th>Binomial Probability:</th>
<th>Geometric Probability:</th>
</tr>
</thead>
<tbody>
<tr>
<td>n =</td>
<td>Y =</td>
</tr>
<tr>
<td>k =</td>
<td>k =</td>
</tr>
<tr>
<td>p =</td>
<td>p =</td>
</tr>
</tbody>
</table>

Mean and Standard Deviation of a Binomial Random Variable

Mean (Expected Value) of a Geometric Random Variable

Identifying distributions:

Examples: Read each scenario involving chance behavior. In each case, determine whether the given random variable has a binomial distribution, a geometric distribution, or neither. Justify your answer.
a) Genetics says that children receive genes from each of their parents independently. Each child of a particular pair of parents has probability 0.25 of having type O blood. Suppose these parents have 5 children. Let $X =$ the number of children with type O blood.

b) Roll a pair of dice until you get doubles. Let $X =$ the number of rolls until you get your first doubles.

c) Shuffle a deck of cards. Turn over the first 10 cards, one at a time. Let $Y =$ the number of aces you observe.

d) Shuffle a deck of cards. Turn over the top card. Put the card back in the deck, and shuffle again. Repeat this process until you get an ace. Let $W =$ the number of cards required.
e) Observer the next 100 cars that go by and let $C = \text{color}$

f) Shoot a basketball 20 times from various distances on the court. Let $Y = \text{number of shots made}$.

CYU: Pg.385

Calculating Binomial Probabilities

Example: Inheriting Blood Type:

Each child of a particular pair of parents has probability 0.25 of having type O blood. Genetics says that children receive genes from each of their parents independently. If these parents have 5 children, the count $X$ of children with type O blood is a binomial random variable with $n =$ and $p =$

Find $P(X = 0)$, $P(X = 1)$, $P(X = 2)$

CYU: Pg.390
Alt. Ex.: Rolling Doubles

In many games involving dice, rolling doubles is desirable. Rolling doubles mean the outcomes of two dice are the same, such as 1&1 or 5&5.

a) The probability of rolling doubles when rolling two dice is ________________.

b) If \( X \) = the number of doubles in 4 rolls of two dice, then \( X \) is ____________________
   with \( n = \) ____ and \( p = \) ________.

c) What is \( P(X = 0) \)? That is, what is the probability that all 4 rolls are not doubles?

d) What is \( P(X = 1) \)?

e) Find the probability that the player gets doubles twice in four attempts.

f) Should the player be surprised if he gets doubles more than twice in four attempts? Justify your answer.

g) Complete the probability distribution and make a histogram of the probability distribution.

<table>
<thead>
<tr>
<th>Value (x)</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability</td>
<td>_____</td>
<td>_____</td>
<td>_____</td>
<td>_____</td>
<td>_____</td>
</tr>
</tbody>
</table>
Example: Bottled Water vs. Tap Water

Mr. Bullard’s AP Statistics class did the water test. There are 21 students in the class. If we assume that the students in his class cannot tell tap water from bottled water, then each one is basically guessing, with a 1/3 chance of being correct. Let \( X \) = the number of students who correctly identify the cup containing the different type of water.

a) Explain why \( X \) is a binomial random variable.

b) Find the mean and standard deviation of \( X \). Interpret each value in context.

c) Of the 21 students in the class, 13 made correct identifications. Are you convinced that Mr. Bullard’s students can tell bottled water from tap water? Justify your answer.

: Pg.393

Binomial Distributions in Statistical Sampling

See Example: Choosing an SRS of CDs (pg.393)

The 10% Condition – Sampling Without Replacement

When taking an SRS of size _____ from a population of size ______, we can use a binomial distribution to model the count of successes in the sample as long as:
When \( n \) is large, we can use Normal probability calculations to approximate binomial probabilities…

### Normal Approximations for Binomial Distributions

Suppose that a count \( X \) has the binomial distribution with \( n \) trials and success probability \( p \). As a rule of thumb, we will use the Normal approximation when \( n \) is so large that:

#### Example: Hiring Discrimination – It Just Won’t Fly! (Sampling without Replacement)

An airline has just finished training 25 first officers – 15 male and 10 female – to become captains. Unfortunately, only eight captain positions are available right now. Airline managers decide to use a lottery to determine which pilots will fill the available positions. Of the 8 captains chosen, 5 are female and 3 are male. Explain why the probability that 5 female pilots are chosen in a fair lottery is not 0.124

#### Alt. Example: Teens and Debit Cards.

In a survey of 506 teenagers ages 14-18, subjects were asked a variety of questions about personal finance. One question asked teens if they had a debit card.

Suppose that exactly 10% of teens ages 14-18 have debit cards. Let \( X \) = the number of teens in a random sample of size 506 that have a debit card.

(a) Show that the distribution of \( X \) is approximately binomial.
(b) Check the conditions for using a Normal approximation in this setting.

(c) Use a Normal distribution to estimate the probability that 40 or fewer teens in the sample have debit cards.

Example: The Birthday Game (Calculating Geometric Probabilities)
See: Activity: The Birth Day Game

The random variable of interest in this game is \( Y = \) number of guesses it takes to correctly match the birth day of one of your teacher’s friends. Each guess is one trial of the chance process.

a) Is this a geometric setting? Justify your answer.

b) Find the probability that the class receives 10 homework problems as a result of playing the game.

c) Find \( P (Y < 10) \) and interpret the value in context.