Bernoulli or Binomial Experiment Properties

- The number of trials in the experiment is fixed.
- There are two outcomes of the experiment:
  - Success
  - Failure
- The probability of success in each trial is the same.
- The trials are independent of each other.
Example: Determine whether each of the following experiments is binomial.

a) A die is rolled 4 times and the uppermost number is observed each time.

b) A marble is drawn and replaced 10 times from a bag containing 6 red and 4 blue marbles.

c) A marble is drawn and not replaced from a bag containing 6 red and 4 blue marbles until there are no marbles left.

d) A coin is tossed repeatedly and the experiment ends after two heads have been observed.

e) A student guesses on every question of a 5-item true-false test.
Example: A multiple-choice exam has 4 questions, each with 5 possible answers. What is the probability of a monkey guessing exactly 3 questions correct?
Computing Binomial Probabilities

Let \( p \) be the probability of success in any trial. Let \( q = 1 - p \) be the probability of failure in any trial. Then in a binomial experiment, the probability of exactly \( r \) successes in \( n \) independent trials is:

\[
P(X = r) = C(n, r) p^r q^{n-r}
\]

- \( X \) is called the *binomial random variable* and the probability distribution of \( X \) is the *binomial distribution*. 
To set up a binomial probability problem, you must do the following:

- Decide that it is a binomial experiment.
  - Check the four criteria.
- Define success.
- Find the number of times the experiment is performed, $n$.
- Find the probability of success, $p$.
- Determine the desired number of successes, $r$. 
**Example:** A new drug being tested causes a serious side effect in 5 out of 100 patients. What is the probability that, in a sample of 10 patients, exactly 2 get the side effect from taking the drug?

*Option 1*
Define success = side effect

\[ n = \text{number of trials} = \]
\[ r = \text{number of successes} = \]
\[ p = \text{probability of success} = \]

*Option 2*
Define success = no side effect

\[ n = \text{number of trials} = \]
\[ r = \text{number of successes} = \]
\[ p = \text{probability of success} = \]
Example: The first half of July was very dry in College Station. If each day there was a 20% chance of rain, what is the probability of no rain in the first 15 days in July?

Define success:

\[ n = \text{number of trials} = \]
\[ r = \text{number of successes} = \]
\[ n - r = \text{number of failures} = \]
\[ p = \text{probability of success} = \]
\[ q = 1 - p = \text{probability of failure} = \]
Example continued: What is the probability of at most 2 days of rain?

\[ r = \text{number of successes} = \]
Example: The probability that a certain calculator component is not working on a given day is 0.04. What is the probability that out of 100 of these components:

a) fewer than 6 are inoperative?

b) more than 3 are inoperative?

c) more than 2 but fewer than 9 are inoperative?
Mean, Variance, and Standard Deviation for Binomial Experiments

- If $X$ is a binomial random variable associated with a binomial experiment consisting of $n$ trials with probability of success $p$ and probability of failure $q = 1 - p$, then
  - Mean: $\mu = E(X) = np$
  - Variance: $\text{Var}(X) = npq$
  - Standard deviation: $\sigma_X = \sqrt{npq}$
Example: Let the random variable $X$ be the number of girls in a 6-child family. Calculate the mean and standard deviation of $X$. 
Example: From our previous example, the probability that a certain calculator component is not working on a given day is 0.04. Calculate the mean, variance, and standard deviation for the number of inoperative components.
Graphs of Probability Distributions for Finite Discrete Variables

- Graph the probability distribution as a histogram.
  - Each rectangle has a base of width 1 (centered on the values of the r. v.) and a height equal to the probability of that value of the r. v.
  - So the area, length $\times$ height, is the probability that a given value of the r. v. occurs.
  - To find the probability of a range of $X$ values, add up the areas over the range of $X$ values.
Continuous Probability Distributions

- Recall that a continuous random variable can take on any real-number value in the appropriate interval.
- A continuous probability distribution is a probability distribution associated with a continuous random variable.
  - Defined by a function $f$ whose domain is the interval of real values taken on by the random variable $X$ associated with the experiment.
Probability Density Function (pdf)

- $f(x)$, the continuous probability distribution, is nonnegative for all values of $x$.
- The area of the region between the graph of $f$ and the $x$-axis is equal to 1.
- $P(a < X < b)$ is the probability that the random variable $X$ assumes a value between $a$ and $b$, $(a < x < b)$.
  - Given by the area of the region between $f(x)$ and the $x$-axis from $x = a$ to $x = b$.
- Graph the probability distribution as a smooth curve.
Normal Distributions

- Many natural, behavioral, and physical phenomena have a continuous distribution with a bell-shaped curve.
- These special families of continuous probability distributions are known as the *normal distributions*. 
Properties of Normal Curves

- The curve has a peak at \( x = \mu \) and is symmetric with respect to the vertical line \( x = \mu \).
- The curve always lies above the \( x \)-axis but approaches the \( x \)-axis as \( x \) extends infinitely in either direction.
- The area under the entire curve is 1.
  - 68.27\% of the area under the curve lies within 1 standard deviation of the mean (that is, between \( \mu - \sigma \) and \( \mu + \sigma \)).
  - 95.45\% of the area lies within 2 standard deviations of the mean.
  - 99.73\% of the area lies within 3 standard deviations of the mean.
- The shape is completely determined by \( \mu \) and \( \sigma \).
The green (medium, centered) curve is the *standard normal distribution* with mean $\mu = 0$ and std. dev. $\sigma = 1$.

- Use $Z$, instead of $X$, as the continuous random variable.
- A small std. dev. yields a tall narrow curve (red) because the data does not spread much from the mean.
- A large std. dev. yields a short wide curve (blue) because the data is spread out.
- The mean shifts the curve left or right. The purple curve has a mean of $-2$. 
Example: On the standard normal curve, what is the probability that a data value is between −1 and 1, i.e. $P(−1 < Z < 1)$?
Example: Suppose that $X$ is a normal random variable with $\mu = 50$ and $\sigma = 10$.

a) What is the probability that $X < 30$?

b) What is the probability that $35 < X < 65$?

c) What is the probability that $X > 70$?
Example: Suppose that $X$ is a normal random variable with $\mu = 50$ and $\sigma = 10$.

a) For what value of $a$ is $P(X < a) = 0.4$?

b) For what value of $a$ is $P(X > a) = 0.6$?

c) For what value of $a$ is $P(X > a) = 0.35$?

 d) For what value of $a$ is $P(\mu - a < X < \mu + a) = 0.5$?
Example: An instructor wants to “curve” the final grades in his class. The averages are normally distributed with a class mean of 73 and a standard deviation of 12. He decides that the top 12% of the class should get an A, the next 24% should get a B, the next 36% a C, the next 18% a D and the last 10% of the class will get an F. What are the cutoffs for the grades?

A cutoff:
B cutoff:
C cutoff:
D cutoff:
Example: A light bulb has an average life of 2000 hours and a standard deviation of 200 hours. The normal distribution closely represents the bulb’s life. Find the probability that a bulb selected at random can be expected to last:

a) between 1900 and 2100 hours.

b) no more than 2200 hours.

c) less than 1500 hours.
Example continued:

d) What bulb life corresponds to the 90th percentile?

e) If a store receives 100-dozen light bulbs, how many would you expect to last at least 2500 hours?
Example: The scores of students on a standardized test form a normal distribution with a mean score of 300 and a standard deviation of 40.

a) What must a student score to be in the upper 10%?

b) In a group of 600 students, how many would be in the upper 10%?

c) What score corresponds to the 70th percentile?
Consider tossing a coin 15 times and counting the number of heads.

- Let $X =$ number of heads.
- This is a binomial experiment with $n = 15, p = 0.5$.
- Since it is binomial, $\mu = 15 \cdot 0.5 = 7.5$ and $\sigma = \sqrt{15 \cdot 0.5 \cdot 0.5} = \sqrt{3.75}$
Look at the normal distribution with $\mu = 7.5$ and $\sigma = \sqrt{3.75}$ and the binomial distribution histogram on the same calculator graph:

- We can use the normal curve to approximate the binomial distribution.
- As the number of coin tosses increases, the binomial distribution comes closer and closer to the normal curve approximation.
Example: Consider tossing a coin 15 times.

a) What is the probability that you toss exactly 5 heads?

b) What is the probability of at least 9 heads?
Example: At a school 1000 children are exposed to the flu. There is a 35% chance of getting the flu if you are exposed. Use the normal curve approximation to the binomial distribution to estimate the probability that:

a) more than 380 children get the flu.

b) fewer than 320 children get the flu.

c) between 320 and 380 children get the flu.