Probability

Experiment is a process that results in an observation that cannot be determined with certainty in advance of the experiment. Each observation is called an outcome or a sample point which may be qualitative or quantitative.

Examples of experiments include (i) Toss a coin (H or T) once, (ii) Toss a coin twice, (iii) through a die (1, 2, 3, 4, 5, 6) once, (iv) Through a die twice, (v) Select $n$ individuals at random from a group of $N$ individuals, (vi) Draw cards from a deck of 52 cards, (vii) Observe a day for rain, no rain (viii) Collect a person’s opinion (for, against) towards a government policy, (ix) Examine a person for the presence or absence of a given disease, etc.

Sample Space ($S$) of an experiment is the set of all possible outcomes or sample points of the experiment. (Recall and compare with the definition of population)

Event ($A$) is a subset of the sample space of an experiment. (Recall and compare with the definition of sample)
Probability of an event $A$

$$P(A) = \text{sum of the probabilities of all sample points that are in } A.$$ 

Probability Rule:

1. Probability of each sample point lies between 0 and 1.
2. Sum of the probabilities of all sample points of a sample space is equal to 1.

Remark:

- *Equally Likely Points* means each point in the sample space has equal probability.
- *Random selection* imply equally likely outcomes.
- *Fair coin* means equal probability of 1/2 for Heads and 1/2 for Tails.
- *Unbiased or Fair Die* means equal probability of 1/6 for each of the six possible points.
- *Proportions* from percentages (often found in journals, newspapers, etc.) are used as probabilities of sample points.]
Example. Two fair coins are tossed, and their up faces (H or T) are recorded. Write down the sample space $S$ and the probabilities of all sample points. Consider the following two events:

   A : observe exactly one head

   B : observe at least one head

Write down the sample points that are in $A$ and $B$. Find the $P(A)$, and $P(B)$.
Union, Intersection, and Complement of Events

Let $S$ be the Sample Space, $A$ and $B$ be two subsets of $S$.

- The union of two events $A$ and $B$, written as $A \cup B$, consists of all points of $S$ that belong to either $A$ or $B$ or both.
- The intersection of $A$ and $B$, written as $A \cap B$, consists of all points of $S$ that belong to both $A$ and $B$.
- Two events $A$ and $B$ are said to be disjoint or mutually exclusive if $A \cap B$ is an empty set.
- The complement of $A$, written as $A^c$, consists of all points of $S$ that are not in $A$.

Laws of Probability

Addition Law : \[ P(A \cup B) = P(A) + P(B) - P(A \cap B) \]

Complementary Law : \[ P(A) + P(A^c) = 1 \]

Example.

Recall the previous coin tossing experiment. Two fair coins are tossed, and their up faces (H or T) are recorded.

$S = \{HH, HT, TH, TT\}$

$A = \{HT, TH\}$

$B = \{HH, HT, TH\}$
Find $A \cup B$, $A \cap B$, $P(A \cup B)$, $P(A \cap B)$, $P(A^c)$.

Example.

The weather at a Caribbean island is perfect for a tourist destination. The chance that it rains on any particular day is only 20% and each day’s weather is independent of every other day’s weather. Suppose you plan a vacation at this island for three days. What is the probability that you get rain on at least one day? one day? no day? at most two days?
Example

The Statistical Abstract of the United States reports that 25% of the country’s households are composed of one person. If 3 randomly selected homes are to participate in a Nielson survey to determine television ratings, find the probability that (i) fewer than two of these homes are one-person households, (ii) at least one home is one-person households.
Example.

The following information is obtained from a group of 1000 students.

Distribution of students according to class rank and employment status

<table>
<thead>
<tr>
<th></th>
<th>Full-time</th>
<th>Part-time</th>
<th>Unemployed</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fr</td>
<td>100</td>
<td>200</td>
<td>10</td>
<td>310</td>
</tr>
<tr>
<td>So</td>
<td>20</td>
<td>80</td>
<td>50</td>
<td>150</td>
</tr>
<tr>
<td>Jr</td>
<td>50</td>
<td>75</td>
<td>25</td>
<td>150</td>
</tr>
<tr>
<td>Sr</td>
<td>70</td>
<td>40</td>
<td>280</td>
<td>390</td>
</tr>
<tr>
<td>Total</td>
<td>240</td>
<td>395</td>
<td>365</td>
<td>1000</td>
</tr>
</tbody>
</table>

If a student is selected at random from these 1000 students, what is the probability
that the student is (i) senior? (ii) unemployed? (iii) senior and unemployed?
Example

The following table summarizes the race and positions of 368 National Basketball Association (NBA) players in 1993 (Sociology of Sport Journal, Vol 14, 1997).

Suppose an NBA player is randomly selected from that year’s player pool.

<table>
<thead>
<tr>
<th></th>
<th>Guard</th>
<th>Forward</th>
<th>Center</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>White</td>
<td>26</td>
<td>30</td>
<td>28</td>
<td>84</td>
</tr>
<tr>
<td>black</td>
<td>128</td>
<td>122</td>
<td>34</td>
<td>284</td>
</tr>
<tr>
<td>Totals</td>
<td>154</td>
<td>152</td>
<td>62</td>
<td>368</td>
</tr>
</tbody>
</table>

(a) What is the probability that the player plays guard but is not African-American?

(b) What is the probability that the player is white or a center?
Example

Let $A$ and $B$ be two subsets of the sample space of an experiment. If $P(A) = 0.4$, $P(B) = 0.5$, and $P(A \cap B) = 0.1$, find (using Venn diagram or otherwise)

(a.) $P(A \cap B^c)$

(b.) $P(A \cup B)^c$

Example

A committee of two students is to be formed at random from a group of five students (three male students M1, M2, M3 and two are female students F1, F2). What is the probability of selecting (i) both male students? (ii) exactly one male student? (iii) at least one female student?

Example

In a factory, machine $A$ produces 25% of all items, machine $B$ 40%, and machine $C$ 35%. Machine $A$ produces 5% defective items, 4% of the items produced by machine $B$ are defective, 3% of the items produced by machine $C$ are defective. What is the probability that a randomly selected item is defective? (Answer: 0.039)
Example

One important finding of a recent study was that 7.8% of children with nonsmoking parents had episodes of pneumonia and/or bronchitis in the first year of life, whereas, 11.4% of children with one smoking parent and 17.6% of children with two smoking parents had such an episode. Suppose that in the general population both parents are smokers in 40% of the households, one parent smokers in 25% of households, and neither parent smokers in 35% of households. What percentage of children in the general population will have pneumonia and/or bronchitis in the first year of life? What is the probability that a randomly selected newborn will have pneumonia and/or bronchitis in the first year of life? (Answer: 0.1262)
• **Conditional Probability** of an event $A$ given that (or knowing that) event $B$ already occurred is defined by

$$P(A/B) = \frac{P(A \cap B)}{P(B)}$$

• **Conditional Probability** of an event $B$ given that (or knowing that) event $A$ already occurred is defined by

$$P(B/A) = \frac{P(A \cap B)}{P(A)}$$

• **Multiplicative Rule of Probability**

$$P(A \cap B) = P(B)P(A/B) = P(A)P(B/A)$$

• **Independent Events.** Two events $A$ and $B$ are independent if and only if $P(A \cap B) = P(A)P(B)$. 
Example.

Two pass-fail exams (one in physics and one in mathematics) are given to each of 200 students. The results are summarized in the following table.

<table>
<thead>
<tr>
<th>Pass in Physics</th>
<th>Fail in Physics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pass in Math</td>
<td>45</td>
</tr>
<tr>
<td>Fail in Math</td>
<td>35</td>
</tr>
</tbody>
</table>

A student is selected at random from 200 students.

a. What is the probability that the student passed math?

b. What is the probability that the student passed math given that the student failed physics?

c. It is known that the student failed in physics. What is the probability that the student passed math?

d. The selected student is found to have passed math. What is the probability that the student failed physics?

e. What is the probability that the student passed both math and physics?

f. Are two events "Fail Math" and "Pass Physics" independent?
Example.

The following information is obtained from a group of 1000 students.

Distribution of students according to class rank and employment status

<table>
<thead>
<tr>
<th></th>
<th>Full-time</th>
<th>Part-time</th>
<th>Unemployed</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fr</td>
<td>100</td>
<td>200</td>
<td>10</td>
<td>310</td>
</tr>
<tr>
<td>So</td>
<td>20</td>
<td>80</td>
<td>50</td>
<td>150</td>
</tr>
<tr>
<td>Jr</td>
<td>50</td>
<td>75</td>
<td>25</td>
<td>150</td>
</tr>
<tr>
<td>Sr</td>
<td>70</td>
<td>40</td>
<td>280</td>
<td>390</td>
</tr>
<tr>
<td>Total</td>
<td>240</td>
<td>395</td>
<td>365</td>
<td>1000</td>
</tr>
</tbody>
</table>

(a) If a student is selected at random from these 1000 students, what is the probability that the student is senior as well as unemployed?

(b) If a student selected at random from these 1000 students is found to be senior, what is the probability that the student is unemployed?
Example.


Suppose an NBA player is randomly selected from that year’s player pool.

<table>
<thead>
<tr>
<th></th>
<th>Guard</th>
<th>Forward</th>
<th>Center</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>White</td>
<td>26</td>
<td>30</td>
<td>28</td>
<td>84</td>
</tr>
<tr>
<td>black</td>
<td>128</td>
<td>122</td>
<td>34</td>
<td>284</td>
</tr>
<tr>
<td>Totals</td>
<td>154</td>
<td>152</td>
<td>62</td>
<td>368</td>
</tr>
</tbody>
</table>

(a) What is the probability that the player plays guard but is not African-American?

(b) What is the probability that the player is white or a center?

(c) What is the probability that the player is a center given that the player is African-American?

(d) Are the events \{white player\} and \{Center\} independent?
Review Problems

1) Answer “TRUE” if the statement is always true. Otherwise answer “False”

—— If $A$ and $B$ are any two events defined on a sample space $S$ of an experiment, then

$$p(A \cap B) = p(A) \cdot p(B)$$

—— If $A$ and $B$ are two disjoint events defined on a sample space $S$ of an experiment, then

$$p(A \cup B) = p(A) + p(B).$$

2) Psychologists tend to believe that there is a relationship between aggressiveness and order of birth. To test this belief, a psychologist chose 500 elementary school students at random and administered each a test designed to measure the student’s aggressiveness. Each student was classified according to one of four categories. The percentages of students falling in the four categories are shown here.

<table>
<thead>
<tr>
<th></th>
<th>Firstborn</th>
<th>Not Firstborn</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aggressive</td>
<td>75</td>
<td>75</td>
</tr>
<tr>
<td>Not Aggressive</td>
<td>125</td>
<td>225</td>
</tr>
</tbody>
</table>

(a.) If one student is chosen at random from the 500, what is the probability that the student is firstborn?

(b.) What is the probability that the student is aggressive, given that the student was firstborn?
(c.) If $A$ : { Student chosen is aggressive }, and $B$ : { Student chosen is firstborn }, are $A$ and $B$ independent? Show your work to support your claim.

2E) Suppose the data in exercise 10 are as follows:

<table>
<thead>
<tr>
<th></th>
<th>Firstborn</th>
<th>Not Firstborn</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aggressive</td>
<td>60</td>
<td>90</td>
</tr>
<tr>
<td>Not Aggressive</td>
<td>140</td>
<td>210</td>
</tr>
</tbody>
</table>

If $A$ : { Student chosen is aggressive }, and $B$ : { Student chosen is firstborn }, are $A$ and $B$ independent? Show your work to support your claim.

3) Investing is a game of chance. Suppose there is a 35% chance that a risky stock investment will end up in a total loss of your investment. Because the rewards are so high, you decide to invest in three independent risky stocks. What is the probability that all three stocks end up in total losses?

4) A human gene carries a certain disease from the mother to the child with a probability rate of 60%. That is, there is a 60% chance that the child becomes infected with the disease. Suppose a female carrier of the gene has four children. Assume that the infections of the four children are independent of one another. Find the probability that at least one child gets the disease from their mother.

MULTIPLE CHOICE. Choose the one alternative that best completes the statement or answers the question.
5) Which of the following probabilities for the sample points A, B, and C could be true if A, B, and C are the only sample points in an experiment?

A) $P(A) = \frac{1}{8}, P(B) = \frac{1}{7}, P(C) = \frac{1}{6}$

B) $P(A) = \frac{1}{4}, P(B) = \frac{1}{4}, P(C) = \frac{1}{4}$

C) $P(A) = -\frac{1}{4}, P(B) = \frac{1}{2}, P(C) = \frac{3}{4}$

D) $P(A) = 0, P(B) = \frac{1}{15}, P(C) = \frac{14}{15}$

6) Fill in the blank. A(n) ———- is the process that leads to a single outcome that cannot be predicted with certainty.

A) sample space  B) sample point  C) event  D) experiment
**Random Variable** is a function that assigns numerical values to sample points (one and only value to each sample point) of the sample space of an experiment. Usual notations are $X$, $Y$, $Z$, etc. A random variable is *discrete* if it can assume only a countable number of values, and *continuous* if it can assume any value in one or more intervals.

**Probability distribution** of a discrete random variable is a graph or a table or a formula that specifies both the values of the variable and the corresponding probabilities. Usually $p(x)$ is used to denote the probability that the random variable $X$ takes the value $x$. Also, note that $p(x) \geq 0$ for all $x$ and $\sum p(x) = 1$.

**Mean** (also called expected value) and **variance** of probability distributions of a discrete random variable:

\[
\text{Mean} = \mu = \sum x p(x)
\]
\[
\text{Variance} = \sigma^2 = \sum x^2 p(x) - \mu^2
\]
Example. Toss a fair coin twice. Let $X$ be the number of heads observed. Then $X$ is a discrete random variable that takes the values 0, 1, 2 with probabilities 1/4, 1/2, and 1/4, respectively. This probability distribution may be displayed as follows.

Probability distribution of $X$ (the number of heads)

<table>
<thead>
<tr>
<th>$X$</th>
<th>$p(x)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1/4</td>
</tr>
<tr>
<td>1</td>
<td>1/2</td>
</tr>
<tr>
<td>2</td>
<td>1/4</td>
</tr>
</tbody>
</table>

Mean and variance:

$$\mu = \Sigma xp(x) = 1$$

$$\text{Variance} = \sigma^2 = \Sigma x^2p(x) - \mu^2 = 1/2$$

The above probability distribution can also be displayed as follows:

$$p(x) = \frac{2!}{x!(2-x)!} \left(\frac{1}{2}\right)^x \left(\frac{1}{2}\right)^{2-x}, \quad x = 0, 1, 2$$
Example.

The number of training units that must be passed before a complex computer software is mastered varies from one to five depending on the student. After much experience, the software manufacturer has determined the following probability distribution that describes the fraction of users mastering the software after each number of training units.

Number of Units : 1 2 3 4 5
Probability of Mastery: .1 .25 .4 .15 .1

a) Calculate the mean number of units necessary to master the program.

b) If the firm wants to ensure that at least 75% of the students master the program, what is the minimum number of training units that must be administered?

Example

The Made Fresh Daily Apple Pie Company knows that the number of pies sold each day varies from day to day. The owner believes that on 50% of the days she sells 100 pies. On another 25% of the days she sells 150 pies, and she sells 200 pies on the remaining 25% of the days. To make sure she has enough product, the owner bakes 200 pies each day at a cost of $2 each. Assume any pies that go unsold are thrown out at the end of the day. If she sells the pies for $4 each, find the probability distribution for her daily profit.
Example. The Showcase Showdown in "The Price is Right".

The game involves a wheel with twenty nickel values 5, 10, ..., 100, marked on it. Contestants spin the wheel once or twice, with the objective of obtaining the highest total score without going over a dollar (100). Let \( x \) represent the outcome for a single contestant playing "The Showcase Showdown". Assume a "fair" wheel. If the total of the player’s spins exceeds 100, the total score is set to zero.

(a) If the player is permitted only one spin of the wheel, find the probability distribution for \( x \).

(b) Suppose the player obtains a 65 on the first spin and decides to spin again. Find the probability distribution for \( x \). What is the probability that the player’s total score exceeds a dollar?

**Binomial Experiment** is consist of \( n \geq 1 \) independent and identical trials where each trial has two possible outcomes \( S \) (for Success) and \( F \) (for Failure) such that \( P(S) = p \) is the same for each of the \( n \) trials. In a binomial experiment, the discrete variable \( X \) (= number of success) is called the binomial random variable and its probability distribution given below is called binomial probability distribution.

\[
p(x) = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}, \quad x = 0, 1, 2, \ldots, n.
\]

Example.
The weather at a Caribbean island is perfect for a tourist destination. The chance that it rains on any particular day is only 20% and each day’s weather is independent of every other day’s weather. Suppose you plan a vacation at this island for ten days. What is the probability that you get rain on two days? one day? no day? at least three days?

Example

The Statistical Abstract of the United States reports that 30% of the country’s households are composed of one person. If 20 randomly selected homes are to participate in a Nielson survey to determine television ratings, find the probability that fewer than five of these homes are one-person households.

Example.

Investing is a game of chance. Suppose there is a 40% chance that a risky stock investment will end up in a total loss of your investment. Because the rewards are so high, you decide to invest in ten independent risky stocks. What is the probability that at least two of these ten stocks end up in total losses?
The probability that a person responds to a mailed questionnaire is 0.4. What is the probability that of 25 questionnaires, more than 15 will be returned?

Example.

The median time a patient waits to see a doctor in a large clinic is 20 minutes. On a day when 25 patients visit the clinic, what is the probability that more than half but fewer than 18 will wait more than 20 minutes?
Continuous Probability Distributions

A function $f$ is called the probability density function of a continuous random variable $X$ if $f$ satisfies the following properties:

i) $f(x) \geq 0$ for all real number $x$.

ii) Total area under the graph of $f$ is 1.

**Definition.**

The probability that $X$ lies between $a$ and $b$ is defined as

$$P(a < X < b) = \text{area between } a \text{ and } b \text{ (under the graph of } f \text{ above the } X\text{-axis)}.$$ 

Note: This course covers only Normal distribution.
Normal Distribution

Example 1. Find each of the following probabilities for the standard normal random variable \( z \)

(a) \( P(-1 \leq z \leq 1) \)

(b) \( P(-1.96 \leq z \leq 1.96) \)

(c) \( P(-2.57 \leq z \leq 2.57) \)

(d) \( P(-1.23 \leq z \leq -0.24) \)

(e) \( P(1.34 \leq z \leq 2.81) \)

(f) \( P(z \leq -1.24) \)

(g) \( P(z \geq 1.85) \)

(h) \( P(z \leq 1.96) \)
Example 2. Find a value of the standard normal random variable $z$, call it $z_0$, such that the following equations hold (you may earn partial credit if you correctly draw the graph and shade the area):

(a) $P(z \leq z_0) = 0.0401$

(b) $P(-z_0 \leq z \leq z_0) = 0.95$

(c) $P(-z_0 \leq z \leq 0) = 0.2967$

(d) $P(z \leq z_0) = 0.0057$

(e) $P(-2 \leq z \leq z_0) = 0.9010$

(f) $P(z \geq z_0) = 0.10$

(g) $P(z \geq z_0) = 0.86$

(h) $P(z \leq z_0) = 0.0405$
Example 3. Suppose the random variable $x$ has a normal distribution with $\mu = 440$ and $\sigma = 50$. Find $x_0$ such that 90% of the values of $x$ are less than $x_0$. 
Example 4. Suppose that the scores, \( x \), on a college entrance examination are normally distributed with a mean score of 540 and a standard deviation of 100.

(a.) What percentage of scores are less than 560?

(b.) What percentage of scores are between 520 and 565?

(c) A certain prestigious university will consider for admission only those applicants whose scores exceed the 90th percentile of the distribution. Find the minimum score an applicant must achieve in order to receive consideration for admission to the university.
Example 5. A physical-fitness association is including the mile run in its secondary-school fitness test for boys. The time for this event for boys in secondary school is approximately normally distributed with a mean of 450 seconds and a standard deviation of 40 seconds. If the association wants to designate the fastest 20% as "excellent," what time should the association set for this criterion?
Example 6. A physical-fitness association is including the mile run in its secondary-school fitness test for boys. The time for this event for boys in secondary school is approximately normally distributed with a mean of 450 seconds and a standard deviation of 40 seconds. What is the probability that six of ten randomly selected boys finish the mile in 400 seconds?
Example 7. Assume that the GPA of all UCF students is normally distributed with mean 2.75 and standard deviation 0.25. Twenty students are randomly selected from all UCF students. What is the probability that eight of these twenty students have GPA exceeding 3.00?
8) Which of the following is not a property of the normal curve?

A) \( P(\mu - 3\sigma < x < \mu + 3\sigma) = 0.997 \)

B) \( P(\mu - \sigma < x < \mu + \sigma) = 0.95 \)

C) mound-shaped (or bell shaped)

D) symmetric about \( \mu \)

9) If a data set is from an approximately normal distribution, approximately what percentage of measurements would you expect to fall within each of the intervals \( \bar{x} \pm s, \bar{x} \pm 2s, \bar{x} \pm 3s \)?

A) 68%, 98%, 100% respectively.

B) 68%, 95%, 100% respectively.

C) 60%, 95%, 98% respectively.

D) 65%, 98%, 100% respectively.
10) The board of examiners that administers the real estate broker’s examination in a certain state found that the mean score on the test was 591 and the standard deviation was 72. If the board wants to set the passing score so that only the best 80% of all applicants pass, what is the passing score? Assume that the scores are normally distributed.
11) The tread life of a particular brand of tire is a random variable best described by a normal distribution with a mean of 60,000 miles and a standard deviation of 6200 miles. If the manufacturer guarantees the tread life of the tires for the first 52,560 miles, what proportion of the tires will need to be replaced under warranty?
TRUE/FALSE.

12) The number of children in a family can be modelled using a continuous random variable.

13) For any continuous probability distribution, \( P(x = c) = 0 \) for all values of \( c \).