PROBABILITY CONCEPTS

1. Relative Frequency as a Intuitive Measure of Probability

Suppose we conduct an experiment for which the outcome cannot be predicted with certainty. An example might a tossing a coin, rolling a die, administering an influenza shot to a 60 year old woman, applying 50 kilograms of nitrogen to an acre of corn, or increasing the money supply by 1%. For each experiment we define some measurement on the outcomes in which we are interested. For example it might be whether the coin lands heads up or tails up, or the amount of nitrogen leaching from the acre of land during the 12 months following the application, or whether the woman contracts a serious (however measured) case of influenza in the six months following injection. In general, suppose we repeat the random experiment a number of times, n, under exactly the same circumstances. If a certain outcome has occurred f times in these n trials; the number f is called the **frequency** of the outcome. The ratio f/n us called the **relative frequency**of the outcome. A relative frequency is typically unstable for small values of n, but it often tends to stabilize about some number, say p, as n increases. The number p is sometimes called the probability of the outcome.

Suppose a researcher randomly selected 100 households in Tulsa, Oklahoma. She then asks each household how many vehicles (truck and cars) they own. The data is below.

Table 1. Data on number of vehicles per household

1	1	2	5	4	5	2	2	4	0	1	1	2	0	0
2	1	1	2	1	1	1	5	2	2	1	6	4	7	1
5	1	3	4	6	2	1	3	1	4	1	0	1	1	2
2	1	6	1	1	3	5	1	1	4	3	4	3	1	0
3	3	0	2	2	3	2	0	6	3	2	4	4	1	0
4	0	1	3	3	6	5	5	2	2	3	2	0	2	4
3	3	2	0	4	1	0	0	1	1					

We can construct a frequency table of the data by listing the all the answers possible (or

Date: October 24, 2005.

received) and the number of times they occurred as follows.

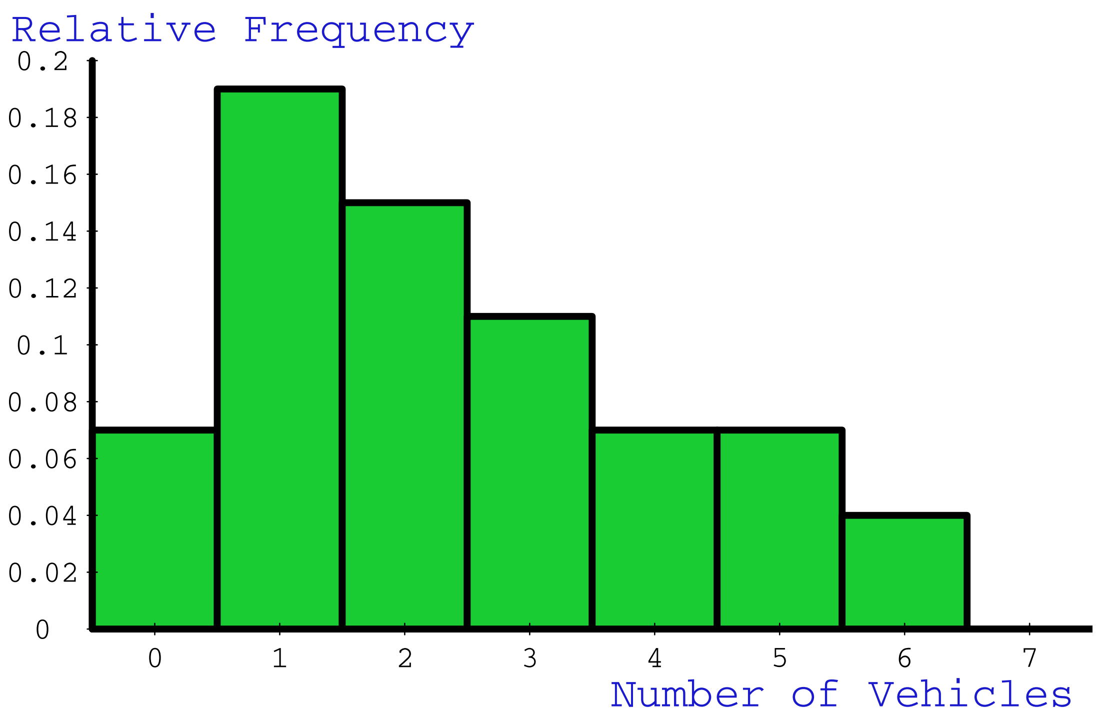
TABLE 2

Number of vehicles	Frequency	Relative Frequency
0	7	0.07
1	19	0.19
2	15	0.15
3	11	0.11
4	7	0.07
5	7	0.07
6	4	0.04
7 or more	0	0.00

One might loosely say from this data that the probability of a household having 4 cars is 7%. Or that the probability of having less than six cars is 96%.

Another way to look at this data is using a histogram. To construct a frequency histogram, we center a rectangle with a base of length one at each observed integer value and make the height equal to the relative frequency of the outcome. The total area of the relative frequency histogram is then one.

FIGURE 1. Histogram.



The value that occurs most often in such frequency data is called the mode and is represented by the rectangle with the greatest height in the histogram.

2. A Review of Set Notation

2.1. **Definition of a Set.** A set is any collection of objects which are called its elements. If x is an element of the set S, we say that x belongs to S and write

$$x \, \varepsilon \, S. \tag{1}$$

If y does not belong to S, we write

$$y \notin S. \tag{2}$$

The simplest way to represent a set is by listing its members. We use the notation

$$A = \{1, \, 2, \, 4\} \tag{3}$$

to denote the set whose elements are the numbers 1, 2 and 4. We can also describe a set by describing its elements instead of listing them. For example we use the notation

$$C = \{x : \; x^2 + 2x - 3 = 0\} \tag{4}$$

to denote the set of all solutions to the equation $x^2 + 2x - 3 = 0$.

2.2. **Subsets.**

2.2.1. *Definition of a Subset.* If all the elements of a set X are also elements of a set Y, then X is a subset of Y and we write

$$X \; \subseteq Y \tag{5}$$

where $\subseteq$ is the set-inclusion relation.

2.2.2. *Definition of a Proper Subset.* If all the elements of a set X are also elements of a set Y, but not all elements of Y are in X, then X is a **proper** subset of Y and we write

$$X \; \subset Y \tag{6}$$

where $\subset$ is the proper set-inclusion relation.

2.2.3. *Definition of Equality of Sets.* Two sets are equal if they contain exactly the same elements, and we write

$$X = Y. \tag{7}$$

2.2.4. *Examples of Sets.*

(i) All corn farmers in Iowa.
(ii) All firms producing steel.
(iii) The set of all consumption bundles that a given consumer can afford,

$$B = \left[(x_1, \, x_2) \, \epsilon \, R^2 : \; x_1 \geq 0, x_2 \geq 0, \; p_1 x_1 + p_2 x_2 \leq I\right] \tag{8}$$

(iv) The set of all combinations of outputs that can be produced by a given set of inputs.

The output set $P(x)$ is the set of all output vectors $y \epsilon R_+^m$ that are obtainable from the input vector $x \epsilon R_+^n$.

$$P(x) = (y \epsilon R_+^m : (x, y) \epsilon T) \tag{9}$$

where the technology set T is the set of all $x \epsilon R_+^n$ and $y \epsilon R_+^m$ such that x will produce y, i.e.,

$$T = \big((x, y) : x \epsilon R_+^n, y \epsilon R_+^m, \text{ such that x will produce y}\big) \tag{10}$$

2.3. **Set Operations.**

2.3.1. *Intersections.* The intersection, W, of two sets X and Y is the set of elements that are in both X and Y. We write

$$W = X \cap Y = \{x : x \epsilon X \text{ and } x \epsilon Y\} \tag{11}$$

2.3.2. *Empty or Null Sets.* The empty set or the null set is the set with no elements. The empty set is written $\emptyset$. For example, if the sets A and B contain no common elements we can write

$$A \cap B = \emptyset \tag{12}$$

and these two sets are said to be disjoint.

2.3.3. *Unions.* The union of two sets A and B is the set of all elements in one or the other of the sets. We write

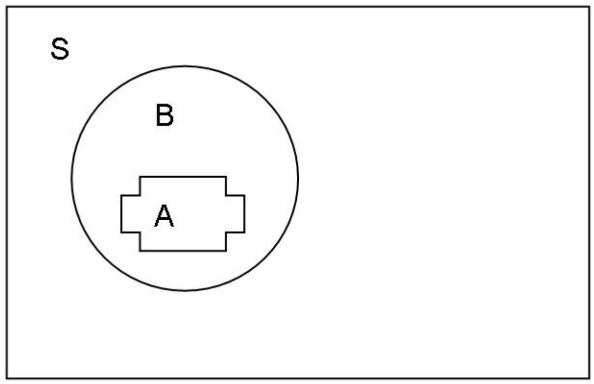
$$C = A \cup B = \{x : x \epsilon A \text{ or } x \epsilon B\}. \tag{13}$$

2.3.4. *Complements.* The complement of a set X is the set of elements of the universal set U that are not elements of X, and is written X^C. Thus

$$X^C = \{x : \epsilon U : x \notin X\} \tag{14}$$

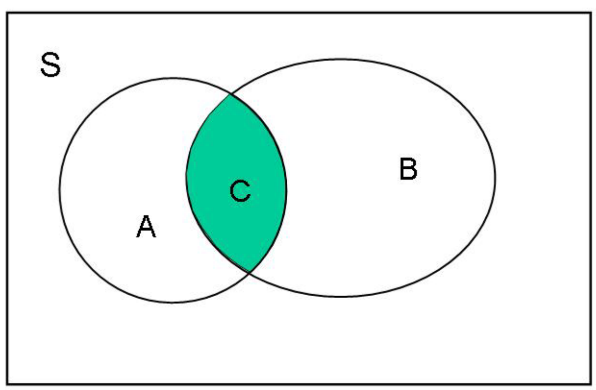
2.3.5. *Venn Diagrams.* A Venn diagram is a graphical representation that is often useful for discussing set relationships. Consider for example figure 2 below. Let S be the universal set, and let A and B be two sets within the universal set.

FIGURE 2



In figure 2, $A \subset B$. Furthermore, $A \cap B = A$ and $A \cup B = B$.
Now consider figure 3.

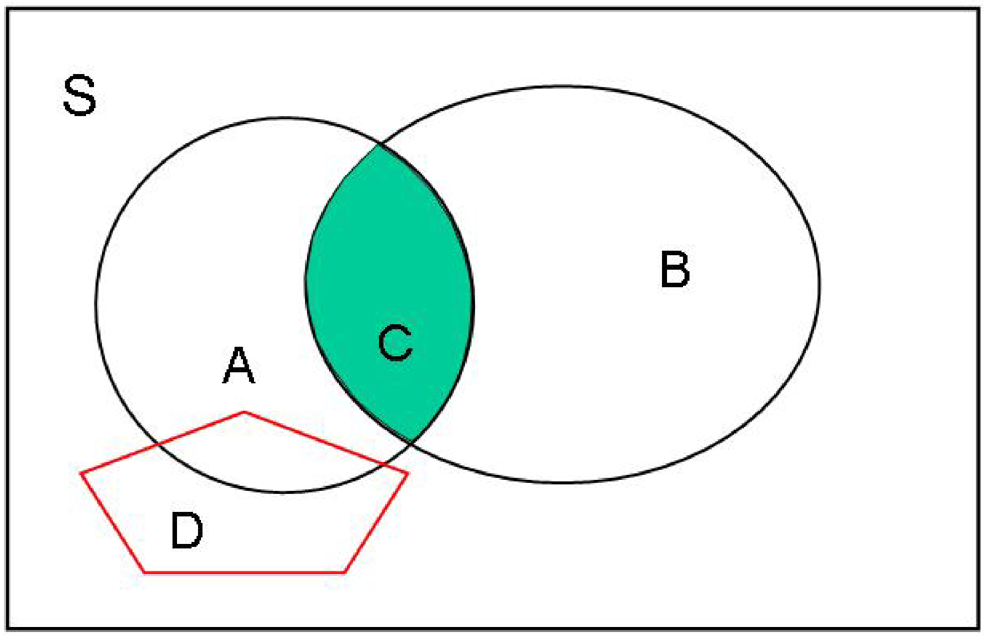
FIGURE 3



In figure 3 the A is the circle to the left and B is the circle to the right. The shaded area marked as C is the area that is shared by both A and B. The union of A and B is everything in both A and B. The intersection of A and B is C. The complement of C in B is the non-shaded portion of B.

Now consider figure 4.

FIGURE 4



In figure 3, $B \cap D = \emptyset$ because they have no points in common. $A \cap D \neq \emptyset$.

2.4. **Distributive Laws.**

(i) $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
(ii) $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$

2.5. **DeMorgan's Laws.**

(i) $(A \cup B)^C = A^C \cap B^C$
(ii) $(A \cap B)^C = A^C \cup B^C$

Consider a proof of the first law. To show that two sets are equal we need to show that:

(i) $\alpha \, \epsilon \, X \Rightarrow \alpha \, \epsilon Y$
(ii) $\alpha \, \epsilon \, Y \Rightarrow \alpha \, \epsilon X$

First prove (i). Suppose that $\alpha \,\epsilon\, (A \cup B)^C$

Then proceed as follows.

$$\begin{aligned} &\alpha \,\epsilon\, (A \cup B)^C \\ &\quad \Rightarrow \alpha \notin A \cup B \\ &\quad \Rightarrow \alpha \notin A \text{ and } \alpha \notin B \\ &\quad \Rightarrow \alpha \,\epsilon\, A \text{ and } \alpha \,\epsilon\, B \\ &\quad \Rightarrow \alpha \,\epsilon\, A^C \cap B^C \end{aligned}$$

Now prove (ii). Suppose that

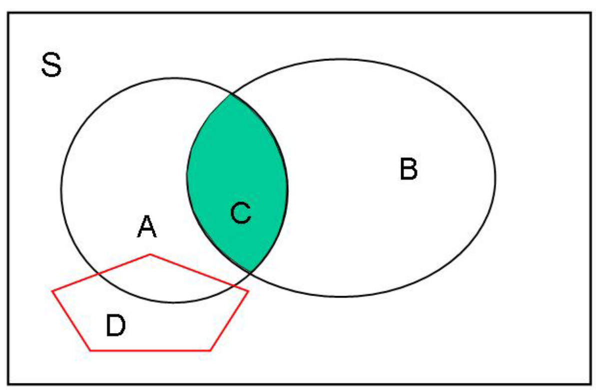
$$\alpha \,\epsilon\, A^C \cap B^C$$

Then proceed as follows.

$$\begin{aligned} &\alpha \,\epsilon\, A^C \cap B^C \\ &\quad \Rightarrow \alpha \,\epsilon\, A^C \text{ and } \alpha \,\epsilon B^C \\ &\quad \Rightarrow \alpha \notin A \text{ and } \alpha \notin B \\ &\quad \Rightarrow \alpha \notin \epsilon\, A \cup B \\ &\quad \Rightarrow \alpha \,\epsilon\, (A \cup B)^C \end{aligned}$$

2.6. **Examples.** Consider the following Venn diagram and see if you can illustrate the following relationships.

FIGURE 5



(i) $D \cap (A \cup B) = (D \cap A) \cup (D \cap B)$
(ii) $D \cup (A \cap E) = (D \cup A) \cap (D \cup E)$
(iii) $(A \cap D)^C = A^C \cup D^C$
(iv) $(A \cup E)^C = A^C \cap E^C$

2.7. **Power Set.** For a given set S, define power set of S,

$$\mathcal{P}_S = A : \; A \subset S = \text{set of all subsets of } S$$

As an example, consider the set $S = \{1,\, 2,\, 3\}$

$$\mathcal{P}_S = \{\phi,\, \{1\},\, \{2\},\, \{3\},\, \{1,2\},\, \{1,3\},\, \{2,3\},\, \{1,2,3\}\}$$

3. Basic Concepts and Definitions of Probability

3.1. **Experiment.** The term **experiment** will denote doing something or observing something happen under certain conditions, resulting in some final state of affairs or "outcome". It might be said that an experiment is the process by which an observation is made. The performance of an experiment is called a **trial** of the experiment. The observed result, on a trial of the experiment, is formally called an **outcome**. The experiment may be physical, biological, social, or otherwise. It may be carried out in a laboratory, in a field trial, or in real-life surroundings, such as an industry, a community, or an economy.

3.2. **Sample Space.** The sample space of an experiment is the set of possible outcomes of the experiment. We denote it by Ω. Its complement, the null or impossible event is denoted by $\emptyset$.

If the experiment consists of tossing a coin, the sample space contains two outcomes, heads and tails; thus,

$$\Omega = \{H,\, T\}$$

If the experiment consists giving a quiz with a maximum score of 10 points to an economics student and only integer scores are recorded, the sample space is

$$\Omega = \{0,\, 1,\, 2,\, 3,\, \ldots\, ,\, 10\}$$

Sample spaces can be either countable or uncountable, if a sample space can be put into a 1–1 correspondence with a subset of the integers, the sample space is countable.

3.3. **Sample Point.** A sample point is any member of Ω and is typically denoted ω.

3.4. **Event.** A subset of Ω is called an event. An event is any collection of possible outcomes of an experiment. A sample point ω is called an elementary or simple event. We can think of a simple event as a set consisting of a single point, the single sample point associated with the event. We denote events by A, B, and so on, or by a description of their members. We say the event A occurs, if the outcome of the experiment is in the set A.

3.5. **Discrete Sample Space.** A discrete sample space is one that contains either a finite or countable number of distinct sample points.

Consider the experiment of tossing a single die. Each time you toss the die, you will observe one and only one simple event. So the single sample point E_1 associated with rolling a one and the single sample point E_5 associated with rolling a five are distinct, and the sets E_1 and E_5 are mutually exclusive. Consider the event denoted A which is rolling an even number. Thus

$$A = \{E_2,\, E_4,\, E_6\}$$

An event in a discrete sample space S is a collection of sample points — that is, any subset of S. Consider for example the event B, where B is the event that the number of the die is four or greater.

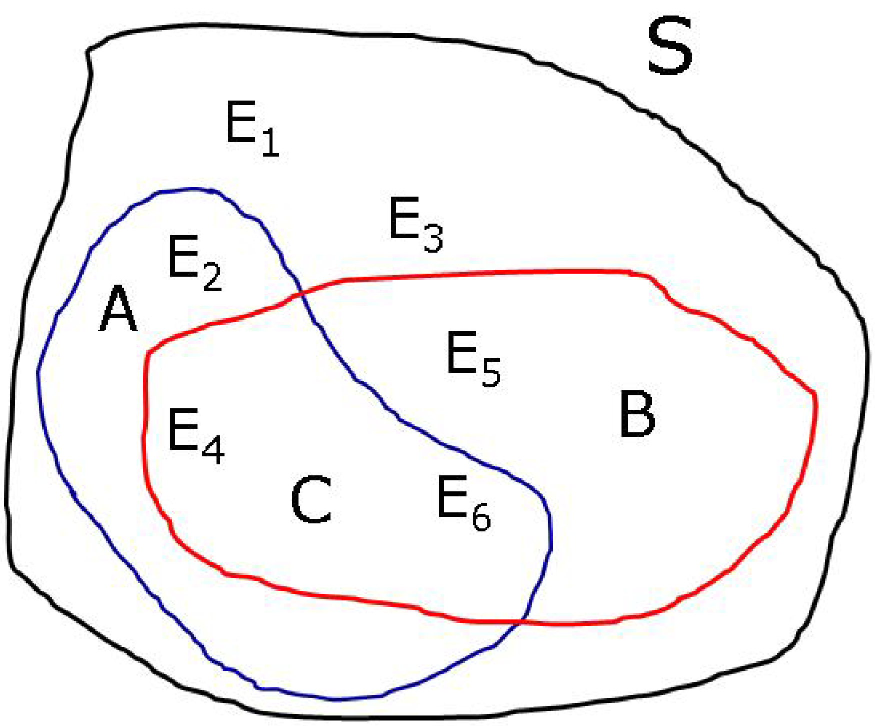
$$B = \{E_4,\, E_5,\, E_6\}$$

Now let C be the intersection of A and B or

$$C = \{E_4,\, E_6\}$$

A Venn diagram might be as follows.

FIGURE 6



3.6. **Sigma Algebra.** A collection of subsets of Ω is called a sigma algebra (or Borel field), denoted by ß , if it satisfies the following three properties:

(i) $\emptyset \,\epsilon$ ß (the empty set is an element of ß)
(ii) If $A \,\epsilon$ ß , then $A^c \,\epsilon$ ß (ß is closed under complementation)

(iii) If $A_1,\ A_2,\ \ldots,\ \epsilon$ ß , then $\bigcup_{i=1}^{\infty} A_i \,\epsilon$ **ß** (ß is closed under countable unions).
Properties (i) and (ii) together that $\Omega\,\epsilon$ ß . Using DeMorgan's laws on intersections and unions we also have
(iv) If $A_1,\ A_2,\ \ldots,\epsilon$ ß , then $\bigcap_{i=1}^{\infty} A_i \,\epsilon$ **ß** (ß is closed under countable intersections.)

3.7. **Examples of Sigma Algebras.**

3.7.1. *Example 1.* If Ω is finite or countable, then ß is easy to define, i.e.

$$ß = \{\text{all subsets of } \Omega \text{ including } \Omega \text{ itself}\}.$$

If Ω has n elements, then there are 2^n sets in ß. For example if $\Omega = \{1,\,2,\,3\}$, then ß is the following collection of sets:

$$\begin{array}{lll} \{1\}\{1,\,2\}\{1,\,2,\,3\} \\ \{2\}\{1,\,3\} & \quad \{\emptyset\} \\ \{3\}\{2,\,3\} \end{array}$$

3.7.2. *Example 2.* Consider a sample space Ω containing four points $\{a,\,b,\,c,\,,d\}$. Consider the collection C consisting of two events: $\{a\}$, $\{c,\,d\}$. If it is desired to expand this collection to a σ-algebra (Borel field), the complements of each event must be included, so $\{b,\,c,\,d\}$ and $\{a,\,b\}$ must be added to the collection. We then have $\{a\}$, $\{c,\,d\}$, $\{b,\,c,\,d\}$ and $\{a,\,b\}$. Because unions and intersections must be included, the events $\{a\,,c,\,d\}$, $\emptyset$, Ω and $\{b\}$ must also be added. In summary, then, a Borel field of sets or events that includes the events $\{a\}$ and $\{c\,,d\}$ consists of the following eight events:

$$\emptyset,\ \{a\},\ \{c,\,d\},\ \{b,\,c,\,d\},\ \{a,\,b\},\ \{a,\,c,\,d\},\ \{b\},\ \Omega$$

Moreover, each of these events is needed, so there is no smaller Borel field that includes the two given events. This smallest one is called the Borel field generated by $\{a\}$ and $\{cd\}$. Observe, though, that the collection of all the $2^4 = 16$ events that can be defined on a sample space with four points is also a Borel field; but it includes the one already defined and is not then the smallest Borel field containing $\{a\}$ and $\{c,\,d\}$.

3.7.3. *Example 3.* Let Ω =(-∞, ∞), the real line. Then we typically choose ß to contain all sets of the form

$$[a,\,b],\ (a,\,b],\ (a,\,b),\ [a,\,b)$$

for all real numbers a and b. From the properties of ß , it is clear that ß contains all sets that can be formed by taking unions and intersections of sets of the above varieties.

3.8. **Probability Defined for a Discrete Sample Space.**

Axiom 1. *The probability of an event is a non-negative number; that is,* $P(A) \geq 0$ *for any subset* A *of* S*.*

Axiom 2. $P(S) = 1$.

Axiom 3. *If* $A_1,\ A_2,\ A_3, \ldots$ *is a finite or infinite sequence of mutually exclusive events of* S*, then*

$$P\left(A_1 \cup A_2 \cup A_3 \cup \cdots\right) = P(A_1) + P(A_2) + P(A_3) + \cdots$$

The second axiom makes sense because one of the possibilities in S must occur. For a discrete sample space we can assign a probability to each simple event satisfying the above axioms to completely characterize the probability distribution. Consider, for example, the experiment of tossing a die. If the die is balanced, it seems reasonable to assign a probability of 1/6 to each of the possible simple outcomes. This assignment agrees with Axiom 1. Axiom 3 indicates that

$$P\left(E_1 \cup E_2 \cup E_3\right) = P(E_1) + P(E_2) + P(E_3)$$

And using Axiom 3, it is clear that Axiom 2 holds.

3.9. **Probability Measure.** Let $\mathcal{F}$ denote a sigma algebra of subsets of Ω, i.e. a class of subsets of Ω to which we assign probabilities. A **probability measure** is a non-negative function P on $\mathcal{F}$ having the following properties:

(i) $P\left(\Omega\right) = 1$.

(ii) If $A_1,\, A_2, \ldots$ are pairwise disjoint sets in $\mathcal{F}$, then

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

We often refer to the triple (Ω, $\mathcal{F}$, P) as a probability model. Only elements of Ω that are members of $\mathcal{F}$ are considered to be events.

3.10. **Elementary Properties of Probability.**

(i) If $A \subset B$, then $P(B - A) = P(B) - P(A)$, where $-$ in the case of sets or events denotes the set theoretic difference.

(ii) $P(A^c) = 1 - P(A),\ P(\emptyset) = 0$.

(iii) If $A \subset B,\ P(B) \geq P(A)$.

(iv) $0 \leq P(A) \leq 1$.

(v) $P\left(\bigcup_{i=1}^{\infty} A_i\right) \leq \sum_{i=1}^{\infty} P\left(A_i\right)$.

(vi) If $A_1 \subset A_2 \subset \ldots A_n,\ \ldots,$ then $P\left(\bigcup_{i=1}^{\infty} A_i\right) = \lim_{n \to \infty} P\left(A_n\right)$.

(vii) $P\left(\bigcup_{i=1}^{k} A_i\right) \geq 1 - \sum_{i=1}^{k} P\left(A_i^c\right)$.

3.11. **Discrete Probability Models.** A probability model is called discrete if Ω is finite or countably infinite and every subset of Ω is assigned a probability. That is, we can write $\Omega = \{\omega_1,\, \omega_2,\, \ldots\}$ and $\mathcal{F}$ is the collection of subsets of Ω. In this case using 3.9 (ii), we can write

$$P\left(A\right) = \sum_{\omega_i \,\epsilon\, A} P\left(\{\omega_i\}\right) \tag{15}$$

We can state this as a theorem.

Theorem 1. *Let $\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\}$ be a finite set. Let $\mathcal{F}$ be any sigma algebra of subsets of Ω. Let $p_1, p_2, \ldots, p_n$ be non-negative numbers such that*

$$\sum_{i=1}^{n} p_i = 1$$

For any $A \in \mathcal{F}$, define $P(A)$ by

$$P(A) = \sum_{(i:\, \omega_i \in A)} p_i. \tag{16}$$

Then P is a probability function on $\mathcal{F}$.

4. Calculating the Probability of an Event Using the Sample Point Method (Discrete Sample Space)

4.1. Steps to Compute Probability.

- (i) Define the experiment and determine how to describe one simple event.
- (ii) List the simple events associated with the experiment and test each one to ensure that it cannot be decomposed.
- (iii) Assign reasonable probabilities to the sample points in S, making certain that $P(E_i) \geq 0$ and $\Sigma P(E_i) = 1$.
- (iv) Define the event of interest, A, as a specific collection of sample points. (A point is in A if A occurs when the sample point occurs. Test all sample points in S to identify those in A.)
- (v) Find $P(A)$ by summing the probabilities of the sample points in A.

4.2. **Example.** A fair coin is tossed three times. What is the probability that exactly two of the three tosses results in heads?

Proceed with the steps as follows.

- (i) The experiment consists of observing what happens on each toss. A simple event is symbolized by a three letter sequence made up of H's and T's. The first letter in the sequence is the result of the first toss and so on.
- (ii) The eight simple events in S are:

$$E_1 : HHH \quad E_2 : HHT \quad E_3 : HTH \quad E_4 : HTT$$
$$E_5 : THH \quad E_6 : THT \quad E_7 : TTH \quad E_8 : TTT$$

- (iii) Given that the coin is fair, we expect that the probability of each event is equally likely; that is,

$$P(E_i) = \frac{1}{8}, \quad i = 1, 2, 3 \ldots$$

- (iv) The event of interest, A, is that exactly two of the tosses result in heads. This implies that

$$A = \{E_2\, E_3\, E_5\}$$

(v) We find $P(A)$ by summing as follows

$$P(A) = P(E_2) + P(E_3) + P(E_5) = \frac{1}{8} + \frac{1}{8} + \frac{1}{8} = \frac{3}{8}$$

5. Tools for Counting Sample Points

5.1. **Probability with Equiprobable Events.** If a sample space contains N equiprobable sample points and an event A contains n_a sample points, then $P(A) = n_a/N$.

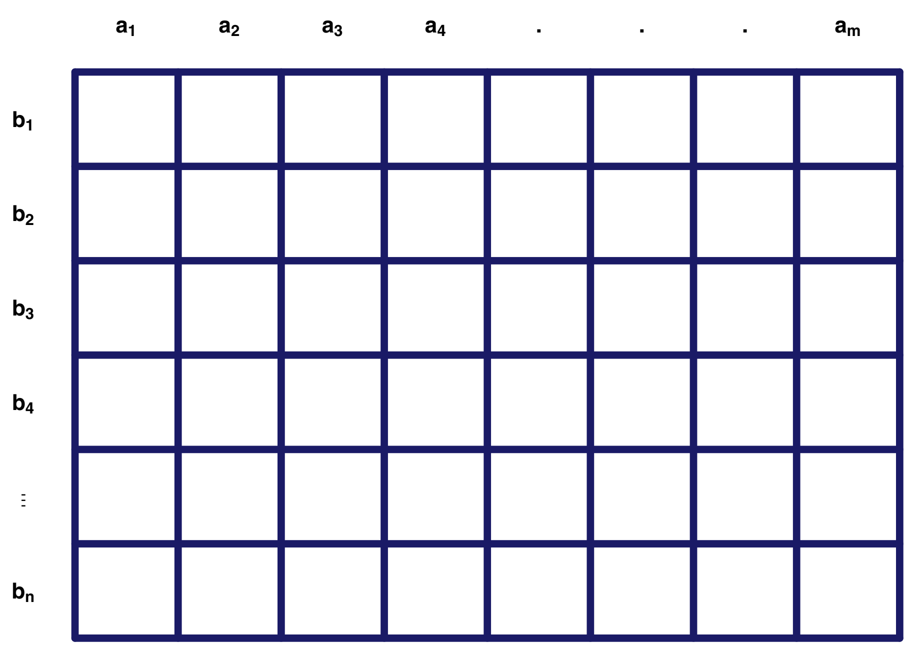
5.2. mn **Rule.**

5.2.1. *Theorem.*

Theorem 2. *With m elements $a_1, a_2, a_3, \ldots, a_m$ and n elements $b_1, b_2, b_3, \ldots, b_n$ it is possible to form $mn = m \times n$ pairs containing one element from each group.*

Proof of the theorem can be can be seen by observing the rectangular table below. There is one square in the table for each a_i, b_j pair and hence a total of $m \times n$ squares.

FIGURE 7. mn rule



5.2.2. The mn rule can be extended to any number of sets. Given three sets of elements, $a_1, a_2, a_3, \ldots, a_m$, $b_1, b_2, b_3 \ldots, , b_n$ and $c_1, c_2, c_3, \ldots, c_\ell$, the number of distinct triplets containing one element from each set is equal to $mn\ell$.

5.2.3. *Example 1.* Consider an experiment where we toss a red die and a green die and observe what comes up on each die.

We can describe the sample space as

$$S_1 = \{(x,\, y) | x = 1,\, 2,\, 3,\, 4,\, 5,\, 6;\; y = 1,\, 2,\, 3,\, 4,\, 5,\, 6\}$$

where x denotes the number turned up on the red die and y represents the number turned up on the green die. The first die can result in six numbers and the second die can result in six numbers so the total number of sample points in S is 36.

5.2.4. *Example 2.* Consider the event of tossing a coin three times. The sets here are identical and consist of two possible elements, H or T. So we have $2 \times 2 \times 2$ possible sample points. Specifically

$$S_2 = \{(x,\, y,\, z) | x = 1,\, 2;\; y = 1,\, 2;\; z = 1,\, 2\}$$

5.2.5. *Example 3.* Consider an experiment that consists of recording the birthday for each of 20 randomly selected persons. Ignoring leap years and assuming that there are only 365 possible distinct birthdays, find the number of points in the sample space S for this experiment. If we assume that each of the possible sets of birthdays is equiprobable, what is the probability that each person in the 20 has a different birthday?

Number the days of the year 1, 2, ..., 365. A sample point for this experiment can be represented by an ordered sequence of 20 numbers, where the first number denotes the number of the day that is the first person's birthday, the second number denotes the number of the day that is the second person's birthday, and so on. We are concerned with the number of twenty-tuples that can be formed, selecting a number representing one of the 365 days in the year from each of 20 sets. The sets are all identical, and each contains 365 elements. Repeated applications of the mn rule tell us there are $(365)^{20}$ such twenty-tuples. Thus the sample space S contains $N = (365)^{20}$ sample points. Although we could not feasibly list all the sample points, if we assume them to be equiprobable, $P(E_i) = 1/(365)^{20}$ for each simple event.

If we denote the event that each person has a different birthday by A, the probability of A can be calculated if we can determine n_a, the number of sample points in A. A sample point is in A if the corresponding 20-tuple is such that no two positions contain the same number. Thus the set of numbers from which the first element in a 20-tuple in A can be selected contains 365 numbers, the set from which the second element can be selected contains 364 numbers (all but the one selected for the first element), the set from which the third can be selected contains 363 (all but the two selected for the first two elements), ..., and the set from which the twentieth element can be selected contains 346 elements (all but those selected for the first 19 elements). An extension of the mn rule yields

$$n_a = (365) \times (364) \times \ldots \times (346).$$

Finally, we may determine that

$$P(A) = \frac{n_a}{N} = \frac{365 \times 364 \times \ldots \times 346}{(365)^{20}} = .5886$$

5.3. **Permutations.**

5.3.1. *Definition.* A ordered arrangement of r distinct objects is called a permutation. The number of ways of ordering n distinct objects taken r at a time is distinguished by the symbol P_r^n

5.3.2. *Theorem.*

Theorem 3.

$$P_r^n = n(n-1)(n-2)(n-3)\,\cdots\,(n-r+1) = \frac{n!}{(n-r)!}$$

5.3.3. *Proof of Theorem.* Apply the extension of the mn rule and note that the first object can be chosen in one of n ways. After the first is chosen, the second can be chosen in $(n-1)$ ways, the third in $(n-2)$ ways, and the

$$P_r^n = n(n-1)(n-2)(n-3)\,\cdots\,(n-r+1)$$

Expressed in terms of factorials, this is given by

$$P_r^n = n(n-1)(n-2)(n-3)\,\cdots\,(n-r+1) = \frac{n!}{(n-r)!}$$

where

$$n! = n(n-1!)(n-2!)(n-3!)\,\cdots\,(2)(1)$$

and

$$0! = 1$$

5.3.4. *Example 1.* Consider a bowl containing six balls with the letters A, B, C, D, E, F on the respective balls. Now consider an experiment where I draw one ball from the bowl and write down its letter and then draw a second ball and write down its letter. The outcome is than an ordered pair. According to theorem 3 the number of distinct ways of doing this is given by

$$P_2^6 = \frac{6!}{4!} = \frac{6\times5\times4\times3\times2\times1}{4\times3\times2\times1} = 6\times5 = 30$$

We can also see this by enumeration.

TABLE 3. Ways of selecting two letters from six in a bowl.

AB	*AC*	*AD*	*AE*	*AF*
BA	*BC*	*BD*	*BE*	*BF*
CA	*CB*	*CD*	*CE*	*CF*
DA	*DB*	*DC*	*DE*	*DF*
EA	*EB*	*EC*	*ED*	*EF*
FA	*FB*	*FC*	*FD*	*FE*

5.3.5. *Example 2.* Suppose that a club consists of 25 members, and that a president and a secretary are to be chosen from the membership. We shall determine the total possible number of ways in which these two positions can be filled. Since the positions can be filled by first choosing one of the 25 members to be president and then choosing one of the remaining 24 members to be secretary, the possible number of choices is

$$P_2^{25} = \frac{n!}{(n-r)!} = \frac{25 \times 24 \times 23 \times \cdots \times 1}{23 \times 22 \times 21 \times \cdots \times 1} = 25 \times 24 = 600$$

5.3.6. *Example 3.* Suppose that six different books are to be arranged on a shelf. The number of possible permutations of the books is

$$P_6^6 = \frac{n}{(n-r)} = \frac{6 \times 5 \times 4 \times \cdots \times 1}{0!} = 6! = 720$$

5.3.7. *Example 4.* Consider the following experiment: A box contains n balls numbered $1, \ldots, n$. First, one ball is selected at random from the box and its number is noted. This ball is then put back in the box and another ball is selected (it is possible that the same ball will be selected again). As many balls as desired can be selected in this way. This process is called *sampling with replacement*. It is assumed that each of the n balls is equally likely to be selected at each stage and that all selections are made independently of each other. Suppose that a total of k selections are to be made, where k is a given positive integer. Then the sample space S of this experiment will contain all vectors of the form $(x_1, \ldots, x_k)$, where x_i is the outcome of the ith selection ($i = l, \ldots, k$). Since there are n possible outcomes for each of the k selections, we can use the mn rule to compute the number of sample points. The total number of vectors in S is n^k. Furthermore, from our assumptions it follows that S is a simple sample space. Hence, the probability assigned to each vector in S is $1/n^k$.

5.4. Partitioning n Distinct Objects into k Distinct Groups.

5.4.1. *Theorem.*

Theorem 4. *The number of ways of partitioning n distinct objects into k distinct groups containing $n_1, n_2, n_3, \ldots, n_k$ objects, respectively, where each object appears in exactly one group and $\Sigma_{i=1}^k n_i = n$, is*

$$N = \binom{n}{n_1 \quad n_2 \quad n_3 \quad \ldots \quad n_k} = \frac{n!}{n_1!\, n_2!\, n_3!\, \ldots\, n_k!}$$

5.4.2. *Proof of Theorem 4.* N is the number of distinct arrangements of n objects in a row for a case in which the rearrangement of the objects within a group does not count. For example, the letters a to ℓ can be arranged into three groups, where $n_1 = 3,\ n_2 = 4$ and $n_3 = 5$:

$$a\,b\,c| \quad d\,e\,f\,g| \quad h\,i\,j\,k\,\ell$$

is one such arrangement. Another arrangement in three groups is

$$a\,b\,d| \quad c\,e\,f\,g| \quad h\,i\ k\,\ell$$

The number of distinct arrangements of n objects, assuming all objects are distinct, is $P_n^n = n!$ from theorem 3. Then P_n^n equals the number of ways of partitioning the n objects into k groups (ignoring order within groups) multiplied by the number of ways of ordering the n_1, n_2, $\ldots$ and n_k elements within each group. This application of the mn rule gives

$$\begin{aligned} P_n^n &= n! = (N)\,(n_1!\, n_2!\, n_3!\, \ldots\, n_k!) \\ \Rightarrow N &= \frac{n!}{n_1!\, n_2!\, n_3!\, \ldots\, n_k!} \equiv \binom{n}{n_1 \quad n_2 \quad n_3 \quad \ldots \quad n_k} \end{aligned}$$

where $n_i!$ is the number of distinct arrangements of n_i objects in group i.

5.4.3. *Example 1.* Consider the letters A, B, C, D. Now consider all the ways of grouping them into two groups of two. Using the formula we obtain

$$\begin{aligned} N &= \frac{n!}{n_1!\, n_2!\, n_3!\, \cdots\, n_{k!}} \\ &= \frac{4!}{2!\, 2!} \\ &= \frac{4 \times 3 \times 2 \times 1}{2 \times 1 \times 2 \times 1} \\ &= 6 \end{aligned}$$

Enumerating we obtain

TABLE 4. Ways of selecting two groups of two letters from four distinct letters.

Group 1	Group 2
AB	*CD*
AC	*BD*
AD	*BC*
BC	*AD*
BD	*AC*
CD	*AB*

5.4.4. *Example 2.* In how many ways can two paintings by Monet, three paintings by Renoir and two paintings by Degas be hung side by side on a museum wall if we do not distinguish between paintings by the same artists. Substituting $n = 7$, $n_1 = 2$, $n_2 = 3$ and $n_3 = 2$ into the formula of theorem 4, we obtain

$$\begin{aligned} N &= \frac{n!}{n_1!\, n_2!\, n_3!} \\ &= \frac{7!}{2!\, 3!\, 2!} \\ &= \frac{7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1}{2 \times 1 \times 3 \times 2 \times 1 \times 2 \times 1} \\ &= 210 \end{aligned}$$

5.5. **Combinations (Arrangement of Symbols Representing Sample Points does not Matter).**

5.5.1. The number of combinations of n objects taken r at a time is the number of subsets, each of size r, that can be formed from the n objects. This number will be denoted by

$$C_r^n \quad \text{or} \quad \binom{n}{r}$$

5.5.2. *Theorem.*

Theorem 5. *The number of unordered subsets of size r chosen (without replacement) from n available objects is*

$$\binom{n}{r} = C_r^n = \frac{P_r^n}{r!} = \frac{n!}{r!\,(n-r)!}$$

5.5.3. *Proof of Theorem 5.* The selection of r objects from a total of n is equivalent to partitioning the n objects into $k = 2$ groups, the r selected and the $(n - r)$ remaining. This is a special case of the general partitioning problem of theorem 4. In the present case, $k = 2$, $n_1 = r$ and $n_2 = (n - r)$. Therefore we have,

$$\binom{n}{r} = \binom{n}{r \;\; n-r} = \frac{n!}{r!(n-r)!}$$

The terms $\binom{n}{r}$ are generally referred to as binomial coefficients because they occur in the binomial expansion

$$\begin{aligned}(x+y)^n &= \binom{n}{0} x^n y^0 + \binom{n}{1} x^{n-1} y^1 + \binom{n}{2} x^{n-2} y^2 + \cdots + \binom{n}{n} x^0 y^0 \\ &= \sum_{i=0}^{n} \binom{n}{i} x^{n-i} y^i\end{aligned}$$

Consider the case of $(x + y)^2$:

$$\begin{aligned}(x+y)^2 &= \binom{2}{0} x^2\, y^0 + \binom{2}{1} x^1 y^1 + \binom{2}{2} x^0 y^2 \\ &= x^2 + 2xy + y^2\end{aligned}$$

Consider the case of $(x + y)^3$:

$$\begin{aligned}(x+y)^3 &= \binom{3}{0} x^3 y^0 + \binom{3}{1} x^2 y^1 + \binom{3}{2} x^1 y^2 + \binom{3}{3} x^1 y^3 \\ &= x^3 + 3x^2 y + 3xy^2 + y^3\end{aligned}$$

5.5.4. *Example 1.* Consider a bowl containing six balls with the letters A, B, C, D, E, F on the respective balls. Now consider an experiment where I draw two balls from the bowl and write down the letter on each of them, not paying any attention to the order in which I draw the balls so that AB is the same as BA. According to theorem 5 the number of distinct ways of doing this is given by

$$C_2^6 = \frac{6!}{2!4!} = \frac{6 \times 5 \times 4 \times 3 \times 2 \times 1}{2 \times 1 \times 4 \times 3 \times 2 \times 1} = \frac{6 \times 5}{2} = 15$$

We can also see this by enumeration.

TABLE 5. Ways of selecting two letters from six in a bowl.

AB	*AC*	*AD*	*AE*	*AF*
	BC	*BD*	*BE*	*BF*
		CD	*CE*	*CF*
			DE	*DF*
				FE

5.5.5. *Example 2.* Suppose that a club consists of 25 members, and that two of them are to be chosen to go to a special meeting with the regional president. We can determine the total possible combinations of individuals who may be chosen to attend the meeting with the formula for a combination.

$$C_2^{25} = \frac{25!}{2!\,23!} = \frac{25 \times 24 \times 23 \times 22 \times \cdots \times 1}{2 \times 1 \times 23 \times 22 \times \cdots \times 1} = \frac{25 \times 24}{2} = 300$$

5.5.6. *Example 3.* Eight politicians meet at a fund raising dinner. How many greetings can be exchanged if each politician shakes hands with every other politician exactly once. Shaking hands with one self does not count. The pairs of handshakes are an unordered sample of size two from a set of eight. We can determine the total possible number of handshakes using the formula for a combination.

$$C_2^{8} = \frac{8!}{2!\,6!} = \frac{8 \times 7 \times 6 \times 5 \times \cdots \times 1}{2 \times 1 \times 6 \times 5 \times \cdots \times 1} = \frac{8 \times 7}{2} = 28$$

6. CONDITIONAL PROBABILITY

6.1. **Definition.** Given an event B such that $P(B) > 0$ and any other event A, we define the conditional probability of A given B, which is written $P(A|B)$, by

$$P\left(A|B\right) = \frac{P(A \cap B)}{P(B)} \tag{17}$$

6.2. **Simple Multiplication Rule.**

$$\begin{aligned} P(A \cap B) &= P(B)P\left(A|B\right) \\ &= P(A)P\left(B|A\right) \end{aligned} \tag{18}$$

6.3. **Consistency of Conditional Probability with Relative Frequency Notion of Probability.** Consistency of the definition with the relative frequency concept of probability can be obtained from the following construction. Suppose that an experiment is repeated a large number, N, of times, resulting in

both A and B,	$A \cap B$, n_{11} times;
A and not B,	$A \cap \bar{B}$, n_{21} times;
B and not A,	$\bar{A} \cap B$, n_{12} times;
and neither A nor B,	$\bar{A} \cap \bar{B}$, n_{22} times.

These results are contained in table 6.

TABLE 6. Frequencies of A and B.

	A	$\bar{A}$	
B	n_{11}	n_{12}	$n_{11}+n_{12}$
$\bar{B}$	n_{21}	n_{22}	$n_{21}+n_{22}$
	$n_{11}+n_{21}$	$n_{12}+n_{22}$	N

Note that $n_{11}+n_{12}+n_{21}+n_{22}=N$.
Then it follows that with large N

$$P(A) \approx \frac{n_{11}+n_{21}}{N}, \quad P(B) \approx \frac{n_{11}+n_{12}}{N}, \quad P(A|B) \approx \frac{n_{11}}{n_{11}+n_{12}} \tag{19}$$

where $\approx$ is read *approximately equal to*.
Now consider the conditional probabilities

$$P(A|B) \approx \frac{n_{11}}{n_{11}+n_{12}}, \quad P(B|A) \approx \frac{n_{11}}{n_{11}+n_{21}}, \quad \text{and} \quad P(A\cap B) \approx \frac{n_{11}}{N} \tag{20}$$

With these probabilities it is easy to see that

$$P(B|A) \approx \frac{P(A\cap B)}{P(A)} \quad \text{and} \quad P(A|B) \approx \frac{P(A\cap B)}{P(B)} \tag{21}$$

6.4. **Examples.**

6.4.1. *Example 1.* Suppose that a balanced die is tossed once. Use the definition to find the probability of a 1, given that an odd number was obtained. Define these events:

***A*:** Observe a 1.
***B*:** Observe an odd number.

We seek the probability of A given that the event B has occurred. The event $A\cap B$ requires the observance of both a 1 and an odd number. In this instance, $A \subset B$ so $A\cap B = A$ and $P(A\cap B) = P(A) = 1/6$. Also, $P(B) = 1/2$ and, using the definition,

$$P(A|B) \approx \frac{P(A\cap B)}{P(B)} = \frac{1/6}{1/2} = \frac{1}{3}$$

6.4.2. *Example 2.* Suppose a box contains r red balls labeled 1, 2, 3,$\ldots$, r and b black balls labeled 1, 2, 3,$\ldots$, b. Assume that the probability of drawing any particular ball is $(b+r)^{-1} = \frac{1}{b+r}$. If a ball from the box is known to be red, what is the probability it is the red ball labeled 1?

First find the probability of A. This is given by

$$P(A) = \frac{r}{b+r}$$

where A is the event that the ball is red. Then compute probability that the ball is the red ball with the number 1 on it. This is given by

$$P(A \cap B) = \frac{1}{b + r}$$

where B is the event that the ball has a 1 on it. Then the probability that the ball is red and labeled 1 given that it is red is given by

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{\frac{1}{b+r}}{\frac{r}{b+r}} = \frac{1}{r}$$

This differs from the probability of B (a 1 on the ball) which is given by

$$P(B) = \frac{2}{b + r}$$

6.4.3. *Example 3.* Suppose that two identical and perfectly balanced coins are tossed once.

(i) What is the conditional probability that both coins show a head given that the first coin shows a head?
(ii) What is the conditional probability that both coins show a head given that at least one of them shows a head?

Let the sample space be given by

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

where each point has a probability of 1/4. Let A be the event the first coin results in a head and B be the event that the second coin results in a head. First find the probabilities of A and B.

$$P(A) = \frac{1}{2}, \quad P(B) = \frac{1}{2}$$

Then find the probability of the intersection and union of A and B.

$$P(A \cap B) = \frac{1}{4}, \quad P(A \cup B) = \frac{3}{4}$$

Now we can find the relevant probabilities. First for (i).

$$P(A \cap B|A) = \frac{P(A \cap B \cap A)}{P(A)} = \frac{P(A \cap B)}{P(A)} = \frac{\frac{1}{4}}{\frac{1}{2}} = \frac{2}{4} = \frac{1}{2}$$

Now for (ii).

$$P(A \cap B|A \cup B) = \frac{P(A \cap B \cap A)}{P(A \cup B)} = \frac{P(A \cap B)}{P(A \cup B)} = \frac{\frac{1}{4}}{\frac{3}{4}} = \frac{1}{3}$$

6.5. **Summation Rule.** If $B_1,\ B_2,\ \ldots\ ,\ B_n$ are (pairwise) disjoint events of positive probability whose union is Ω, the identity

$$A = \bigcup_{j=1}^{n} (A \cap B_j)$$

??, and (18) yield

$$P\,(A) = \sum_{j=1}^{n} P(A|B_j)P(B_j), \tag{22}$$

Consider rolling a die. Let the sample space be given by

$$\Omega = \{1,\, 2,\, 3,\, 4,\, 5,\, 6\}$$

Now consider a partition B_1, B_2, B_3, given by

$$B_1 = \{1,\, 2\} \quad B_2 = \{3,\, 4\} \quad B_3 = \{5,\, 6\}$$

Now let A be the event that the number on the die is equal to one or greater than or equal to four. This gives

$$A = \{1,\, 4,\, 5,\, 6\}$$

The probability of A is $4/6 = 2/3$. We can also compute this using the formula in (22) as follows

$$\begin{aligned} P\,(A) &= P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + P(A|B_3)P(B_3) \\ &= \left(\frac{1}{2}\right)\left(\frac{1}{3}\right) + \left(\frac{1}{2}\right)\left(\frac{1}{3}\right) + (1)\left(\frac{1}{3}\right) \\ &= \frac{1}{6} + \frac{1}{6} + \frac{1}{3} \\ &= \frac{2}{3} \end{aligned} \tag{23}$$

6.6. **Bayes Rule.**

$$P\,(B_i|A) = \frac{P\,(A|B_i)\,P(B_i)}{\sum_{j=1}^{n} P(A|B_j)P(B_j)} \tag{24}$$

Consider the same example as in 6.5. First find the probability for B_1.

$$P(B_1|A) = \frac{P(A|B_1)\,P(B_1)}{\sum_{j=1}^{n} P(A|B_j)P(B_j)} = \frac{\left(\frac{1}{2}\right)\left(\frac{1}{3}\right)}{\frac{2}{3}} = \frac{\frac{1}{6}}{\frac{2}{3}} = \frac{1}{4} \tag{25}$$

Now for B_2.

$$P(B_2|A) = \frac{P(A|B_2)\,P(B_2)}{\sum_{j=1}^{n} P(A|B_j)P(B_j)} = \frac{\left(\frac{1}{2}\right)\left(\frac{1}{3}\right)}{\frac{2}{3}} = \frac{\frac{1}{6}}{\frac{2}{3}} = \frac{1}{4} \tag{26}$$

And finally for B_3.

$$P(B_3|A) = \frac{P(A|B_3)\,P(B_3)}{\sum_{j=1}^{n} P(A|B_j)P(B_j)} = \frac{(1)\left(\frac{1}{3}\right)}{\frac{2}{3}} = \frac{\frac{1}{3}}{\frac{2}{3}} = \frac{1}{2} \tag{27}$$

6.7. **Conditional Probability Given Multiple Events.** The conditional probability of A given $B_1, B_2, \ldots, B_n$, is written $P(A|B_1,\ B_2,\ \ldots\ , B_n)$ and is defined by

$$P(A|B_1,\ B_2,\ \ldots\ ,\ B_n) = P(A|B_1 \cap B_2 \cap B_3\ \cdots \cap B_n) \tag{28}$$

for any events A, $B_1, B_2, \ldots, B_n$ such that $P(B_1 \cap B_2 \cap \cdots \cap B_n) > 0$.

6.8. **More Complex Multiplication Rule.**

$$P(B_1 \cap B_2 \cap B_3 \cdots \cap B_n) = P(B_1)P(B_2|B_1)P(B_3|B_1B_2)\cdots P(B_n|B_1,\ B_2, \cdots B_{n-1}) \tag{29}$$

7. Independence

7.1. **Definition.** Two events are said to be independent if

$$P(A \cap B) = P(A)P(B) \tag{30}$$

If $P(B) > 0$ (or $P(A) > 0$), this can be written in terms of conditional probability as

$$\begin{aligned} P(A|B) &= P(A) \\ P(B|A) &= P(B) \end{aligned} \tag{31}$$

The events A and B are independent if knowledge of B does not affect the probability of A.

7.2. **Examples.**

7.2.1. *Example 1.* Suppose that two identical and perfectly balanced coins are tossed once. Let the sample space be given by

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

where each point has a probability of 1/4. Let A be the event the first coin results in a head, B be the event that the second coin results in a tail, and C the event that both flips in tails. First find the probabilities of A, B, and C.

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

Now find the probability of the intersections of A, B and C.

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

Now find the probability of B given A and B given C.

First find the probability of B given A

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{\frac{1}{4}}{\frac{2}{4}} = \frac{1}{2}$$

This is the same as $P(B)$ so A and B are independent. Now find

$$P(B|C) = \frac{P(B \cap C)}{P(C)} = \frac{\frac{1}{4}}{\frac{1}{4}} = 1$$

7.2.2. *Example 2.* Roll a red die and a green die. Let $A = \{4 \text{ on the red die})$ and $B = \{\text{sum of the dice is odd}\}$. Find $P(A)$, $P(B)$, and $P(A \cap B)$. Are A and B independent?

There are 36 points in the sample space.

TABLE 7. Possible outcomes of rolling a red die and a green die. First number in pair is number on red die.

Green (A) Red (D)	1	2	3	4	5	6
1	1 1	1 2	1 3	1 4	1 5	1 6
2	2 1	2 2	2 3	2 4	2 5	2 6
3	3 1	3 2	3 3	3 4	3 5	3 6
4	4 1	4 2	4 3	4 4	4 5	4 6
5	5 1	5 2	5 3	5 4	5 5	5 6
6	6 1	6 2	6 3	6 4	6 5	6 6

The probability of A is $6/36 = 1/6$. The probability of B is $18/36 = 1/2$ and the probability of $A \cap B$ is $3/36 = 1/12$. To check for independence multiply $P(A)$ and $P(B)$ as follows

$$P(A)P(B) = \left(\frac{1}{6}\right)\left(\frac{1}{2}\right) = \frac{1}{12}$$

The events A and B are thus independent.

7.2.3. *Example 3.* For example 2 compute $P(A|B)$. Are A and B independent?

There are 18 sample points associated with B. Of these three belong to A so that

$$P(A|B) = \frac{3}{18} = \frac{1}{6}$$

Using the relationship that

$$P(A|B) = P(A)$$

under independence, we can see that A and B are independent.

7.2.4. *Example 4.* Now consider the same experiment. Let $C = \{5 \text{ on the red die})$ and $D = \{\text{sum of the dice is eleven}\}$.

Find $P(C)$, $P(D)$ and $P(C \cap D)$. Are C and D independent? Can you show the same thing using conditional probabilities?

The probability of C is $6/36 = 1/6$. The probability of D is $2/36 = 1/18$ and the probability of $C \cap D$ is $1/36$. To check for independence multiply $P(C)$ and $P(D)$ as follows:

$$P(C)P(D) = \left(\frac{1}{6}\right)\left(\frac{1}{18}\right) = \frac{1}{108}$$

The events A and B are thus not independent.

To show dependance using conditional probability, compute the of conditional probability C given D.

There are two ways that D can occur. Given that D has occurred, there is one way that C can occur so that

$$P(C|D) = \frac{1}{2}$$

This is not equal to the probability of C so C and D are not independent.

7.2.5. *Example 5.* Die A has orange on one face and blue on five faces; die B has orange on two faces and blue on four faces; die C has orange on three faces and blue on three faces. The dice are fair. If the three dice are rolled, what is the probability that exactly two of the dice come up orange.

There are three ways that two of the dice can come up orange: AB, AC and BC. Because the events are mutually exclusive we can compute the probability of each and then add them up. Consider first the probability that A is orange, B is orange and C is not orange. This is given by

$$P(AB) = \left(\frac{1}{6}\right)\left(\frac{2}{6}\right)\left(\frac{3}{6}\right) = \frac{6}{216}$$

because the chance of an orange face on A is $1/6$, the chance of an orange face on B is $2/6$ and the chance of blue on C is $3/6$ and the events are independent.

Now consider the probability that A is orange, B is blue and C is orange. This is given by

$$P(AC) = \left(\frac{1}{6}\right)\left(\frac{4}{6}\right)\left(\frac{3}{6}\right) = \frac{12}{216}$$

because the chance of an orange face on A is 1/6, the chance of an blue face on B is 4/6 and the chance of orange on C is 3/6 and the events are independent.

Now consider the probability that A is blue, B is orange and C is orange. This is given by

$$P(BC) = \left(\frac{5}{6}\right)\left(\frac{2}{6}\right)\left(\frac{3}{6}\right) = \frac{30}{216}$$

because the chance of an blue face on A is 5/6, the chance of an orange face on B is 2/6 and the chance of orange on C is 3/6 and the events are independent. Now add these up to obtain the probability of two orange dice.

$$P(\text{two orange dice}) = \frac{6}{216} + \frac{12}{216} + \frac{30}{216} = \frac{48}{216} = \frac{2}{9}$$

7.3. **More General Definition of Independence.** The events $A_1,\ A_2,\ \ldots\ ,\ A_n$ are said to be independent if

$$P\left(A_{i_1} \cap A_{i_2} \cdots \cap A_{i_k}\right) = \prod_{j=1}^{k} P\left(A_{i_j}\right) \tag{32}$$

for any subset $\{i_1, i_2,\ \ldots\ ,\ i_k\}$ of the integers $\{1, 2,\ \ldots\ ,\ , n\}$. If all the $P(A_i)$ are positive, we can rewrite this in terms of conditional probability as

$$P\left(A_j | A_{i_1},\ A_{i_2},\ \ldots\ ,\ A_{i_k}\right) = P\left(A_j\right) \tag{33}$$

for any j and $\{i_1,\ i_2,\ \ldots\ ,\ i_k\}$ such that $j \notin \{i_1,\ i_2,\ \ldots\ ,\ i_k)$.

7.4. **Additive Law of Probability.** The probability of the union of two events A and B is given by

$$P\left(A \cup B\right) = P(A) + P(B) - P(A \cap B) \tag{34}$$

If A and B are mutually exclusive events, $P(A \cap B) = \emptyset$ and

$$P\left(A \cup B\right) = P(A) + P(B) \tag{35}$$

which is the same as Axiom 3 for probability defined for a discrete sample space or property two of a probability measure defined on a sigma algebra of subsets of a sample space Ω. Similarly for three events, A, B and C we find

$$\begin{aligned} &P\left(A \cup B \cup C\right) \\ &\quad = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C) \end{aligned} \tag{36}$$

8. Random Variables

8.1. **Somewhat Intuitive Definition of a Random Variable.** Consider an experiment with a sample space Ω. Then consider a function X that assigns to each possible outcome in an experiment ($\omega \in \Omega$) one and only one real number, $X(\omega) = x$. The space of X is the set of real numbers $\{x : X(\omega) = x, \omega \in \Omega\}$, where $\omega \in \Omega$ means that the element ω belongs to the set Ω. If X may assume any value in some given interval I (the interval may be bounded or unbounded), it is called a continuous random variable. If it can assume only a number of separated values, it is called a discrete random variable.

An even simpler (but less precise) definition of a random variable is as follows. A random variable is a real-valued function for which the domain is a sample space.

Alternatively one can think of a random variable as a set function, it assigns a real number to a set. Consider the following example.

Roll a red die and a green die. Let the random variable assigned to each outcome be the sum of the numbers on the dice. There are 36 points in the sample space. In table 8 the outcomes are listed along with the value of the random variable associated with each outcome.

Table 8. Possible outcomes of rolling a red die and a green die. First number in pair is number on red die.

Green (A) / Red (D)	1	2	3	4	5	6
1	1 1 *2*	1 2 *3*	1 3 *4*	1 4 *5*	1 5 *6*	1 6 *7*
2	2 1 *3*	2 2 *4*	2 3 *5*	2 4 *6*	2 5 *7*	2 6 *8*
3	3 1 *4*	3 2 *5*	3 3 *6*	3 4 *7*	3 5 *8*	3 6 *9*
4	4 1 *5*	4 2 *6*	4 3 *7*	4 4 *8*	4 5 *9*	4 6 *10*
5	5 1 *6*	5 2 *7*	5 3 *8*	5 4 *9*	5 5 *10*	5 6 *11*
6	6 1 *7*	6 2 *8*	6 3 *9*	6 4 *10*	6 5 *11*	6 6 *12*

We then write $X = 8$ for the subset of Ω given by

$$\{(6,\, 2)(5,\, 3)(4,\, 4)(3,\, 5)(2,\, 6)\}$$

Thus, $X = 8$ is to be interpreted as the set of elements of Ω for which the total is 8.

Or consider another example where a student has 5 pairs of brown socks and 6 pairs of black socks or 22 total socks. He does not fold his socks but throws them all in a drawer.

Each morning he draws two socks from his drawer. What is the probability he will choose two black socks? We will first develop the sample space where we use the letter B to denote black and the letter R (for russet) to denote brown. There are four possible outcomes to this experiment: BB, BR, RB and RR. Let the random variable defined on this sample space be the number of black socks drawn.

TABLE 9. Sample space, values of random variable and probabilities for sock experiment

Element of Sample Space	**Value of Random Variable**	**Probability of Random Variable**
BB	2	$\left(\frac{12}{22}\right)\left(\frac{11}{21}\right) = \frac{22}{77}$
BR	1	$\left(\frac{12}{22}\right)\left(\frac{10}{21}\right) = \frac{20}{77}$
RB	1	$\left(\frac{10}{22}\right)\left(\frac{12}{21}\right) = \frac{20}{77}$
RR	0	$\left(\frac{10}{22}\right)\left(\frac{9}{21}\right) = \frac{15}{77}$

We can compute the probability of two black socks by multiplying 12/22 by 11/21 to obtain 22/77. We can compute the other probabilities in a similar manner.

8.2. **Note on Borel Fields and Open Rectangles.**

8.2.1. *Open Rectangles.* If $(a_1,\, b_1),\, (a_2,\, b_2),\, \ldots\, ,\, (a_k,\, b_k)$ are k open intervals, then the set

$$(a_1,\, b_1)\, \times\, (a_2,\, b_2)\, \times\, \cdots \times (a_k,\, b_k) = \{(x_1,\, x_2,\, \ldots\, ,\, x_k) : a_i < x_i < b_i,\, 1 \leq I \leq k\}$$

is called an open k rectangle. If $k = 1$ then we have an open interval on the real line, if $k = 2$, we have a rectangle in R^2, and so on.

8.2.2. *Borel Fields in R^k.* The smallest σ-field having all the open rectangles in R^k as members is called the Borel field in R^k and is denoted ß^k. ß^1 is the smallest σ-field that contains all the open intervals of the real line.

8.3. **Formal Definition of a Random Variable.** A random variable X is a function from the sample space Ω to the real line R such that the set

$$\{\omega :\, X(\omega)\, \epsilon\, B\} = X^{-1}(B)$$

is in $\mathcal{F}$ (the sigma field in our probability model) for every B that is an element of ß^1. The idea (loosely) is that we pick some interval of the real line and insist that when we map this interval back to the sample space Ω, the elements of Ω so obtained are in $\mathcal{F}$. In technical terms this means that the function X must be measurable. Formally, a function $X(\omega)$ defined on a sample space is called measurable with respect to $\mathcal{F}$ if and only if the event $\{\omega | X(\omega) \leq \lambda\}$ is in $\mathcal{F}$ for every real λ. The significance of this condition is that when probabilities have been assigned to events in Ω, the event $X(\omega) \leq \lambda$ will have a probability, because it is one of the sets in $\mathcal{F}$.

Consider the example where a red die and a green die are rolled and the random variable is the sum of the numbers on the dice. In this case X^{-1}(9) is the set

$$\{(6,\,3)(5,\,4)(4,\,5)(3,\,6)\}$$

8.4. **Notation.** We will normally use *uppercase letters*, such as X to denote *random variables*, and *lowercase letters*, such as x, to denote particular values that a random variable may assume. For example, let X denote any one of the six possible values that can result from tossing a die. After the die is tossed, the number actually observed will be denoted by the symbol x. Note that X is a random variable, but the specific observed value, s, is not random. We sometimes write $P(X = x)$ for the probability that the random variable X takes on the value x.

8.5. **Random Vector.** A random vector $X = (X_1,\, X_2,\, \ldots\,,\, X_k)'$ is a k-tuple of random variables. Formally it is a function from Ω to R^k such that the set

$$\{\omega :\; X(\omega)\,\epsilon\, B\} = X^{-1}(B)$$

is in $\mathcal{F}$ for every B that is an element of ß^k. For $k = 1$, random vectors are just random variables. The event $X^{-1}(B)$ (meaning the event that leads to the function X mapping it into the set B in R^k) is usually written $[X\,\epsilon\, B]$. This means that when we write $[X\,\epsilon\, B]$ we mean that the event that occurred lead to an outcome that mapped into the set B in R^k. And when we write $P[X\,\epsilon\, B]$, we mean the probability of the event that will map into the set B in R^k.

8.6. **Probability Distribution of a Random Vector.** The probability distribution of a random vector S is defined as the probability measure in the probability model (R^k, ß^k, P_X). This is given by

$$P_X(B) = P([X\,\epsilon\, B]) \tag{37}$$

9. Random Sampling

9.1. **Simple Random Sampling.** Simple random sampling is the basic sampling technique where we select a group of subjects (a sample) for study from a larger group (a population). Each individual is chosen entirely by chance and each member of the population has an equal chance of being included in the sample. Every possible sample of a given size has the same chance of selection, i.e. each member of the population is equally likely to be chosen at any stage in the sampling process.

If, for example, numbered pieces of cardboard are drawn from a hat, it is important that they be thoroughly mixed, that they be identical in every respect except for the number printed on them and that the person selecting them be well blindfolded.

Second in order to meet the equal opportunity requirement, it is important that the sampling be done with replacement. That is, each time an item is selected, the relevant measure is taken and recorded. Then the item must be replaced in the population and be thoroughly mixed with the other items before the next item is drawn. If the items are not replaced in the population, each time an item is withdrawn, the probability of being selected, for each of the remaining items, will have been increased. For example, if in the black sock brown sock problem, we drew two black socks from the drawer and put them on, and then repeated the experiment, the probability of two black socks would now be $10/20$ multiplied by $9/19$ rather than $12/22 \times 11/21$. Of course, this kind of change in probability becomes trivial if the population is very large.

More formally consider a population with N elements and a sample of size n. If the sampling is conducted in such a way that each of the $\binom{N}{n}$ samples has an equal probability of bring selected, the sampling is said to be random and the result is said to be a simple random sample.

9.2. **Random Sampling.** Random sampling is a sampling technique where we select a group of subjects (a sample) for study from a larger group (a population). Each individual is chosen entirely by chance and each member of the population has a known, but possibly non-equal, chance of being included in the sample.

REFERENCES

[1] Amemiya, T. *Advanced Econometrics.* Cambridge: Harvard University Press, 1985.
[2] Bickel, P. J. and K. A. Doksum. *Mathematical Statistics: Basic Ideas and Selected Topics, Vol 1.* 2^{nd} Edition. Upper Saddle River, NJ: Prentice Hall, 2001.
[3] Billingsley, P. *Probability and Measure.* 3rd edition. New York: Wiley, 1995.
[4] Casella, G. and R. L. Berger. *Statistical Inference*. Pacific Grove, CA: Duxbury, 2002.
[5] Cramer, H. *Mathematical Methods of Statistics.* Princeton: Princeton University Press, 1946.
[6] Goldberger, A. S. *Econometric Theory.* New York: Wiley, 1964.
[7] Lindgren, B. W. *Statistical Theory* 3^{rd} edition. New York: Macmillan Publishing Company, 1976.
[8] Rao, C. R. *Linear Statistical Inference and its Applications.* 2nd edition. New York: Wiley, 1973.