Probability and Statistics (ENM 503)

Michael A. Carchidi

October 8, 2018

Chapter 3 - Introduction to Sets and Probability

The following notes are based on the textbook entitled: *A First Course in Probability* by Sheldon Ross (9th edition) and these notes can be viewed at

https://canvas.upenn.edu/

after you log in using your PennKey user name and Password.

1. A Review of Set Terminology and Definitions

In the study of probability, the mathematics of sets and operations defined on sets is a valuable tool. Therefore it is natural to begin our study of probability with a review of sets and some properties of sets. If you are already familiar with sets, you should still quickly read this section just to see the notation that we shall be using throughout the rest of the course.

A *set* is simply a collection of objects (called *elements* of the set). We shall denote sets using upper-cased italic letters $(A, B, C, \ldots)$ and an object a is an element of a set A (written as $a \in A$) if the object a belongs to the set A. If an object a does not belong to the set A, we write $a \notin A$. A set A is called *finite* if it contains a finite number of elements and its called *infinite* otherwise. If A is finite, then we denote the number of elements in A (or the *cardinality of* A) by $|A|$. Thus if $A = \{1, 2, 3, 4, 5, 6\}$, then $|A| = 6$.

The Empty Set and Universal Set

The set that contains no objects is called the *empty set* and is denoted by the symbol $\emptyset$, and hence $|\emptyset| = 0$. Note then that for all objects a, we have $a \notin \emptyset$. A set that contains all objects of interest in an experiment is called the *sample set (or universal set) of interest* and is denoted by S, and the sample set can change depending on the experiment and depending on what is recorded in an experiment.

For example, if an experiment consists of flipping a coin and recording the outcome, heads (H) or tails (T), then $S = \{H, T\}$, but if the experiment consists of rolling a die and recording the outcome, then $S = \{1, 2, 3, 4, 5, 6\}$. If a experiment consists of rolling two dice and recording the outcome of both die, then the sample space would be

$$\begin{aligned} S = \{&(1,1),(1,2),(1,3),(1,4),(1,5),(1,6),\\ &(2,1),(2,2),(2,3),(2,4),(2,5),(2,6),\\ &(3,1),(3,2),(3,3),(3,4),(3,5),(3,6),\\ &(4,1),(4,2),(4,3),(4,4),(4,5),(4,6),\\ &(5,1),(5,2),(5,3),(5,4),(5,5),(5,6),\\ &(6,1),(6,2),(6,3),(6,4),(6,5),(6,6)\}. \end{aligned}$$

However, for the same experiment of rolling two dice, we might be interested in recording only the sum of the two die, which would then result in a sample space given as

$$S = \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}.$$

The Compliment of a Set and the Subset of a Set

The *compliment* of a set A in S (denoted by $\bar{A}$) is the set of all objects in S that are not in A, which we also write as

$$\bar{A} = S - A. \tag{1}$$

Note that

$$\bar{\bar{A}} = A \qquad , \qquad \bar{\emptyset} = S \qquad \text{and} \qquad \bar{S} = \emptyset \tag{2}$$

and if S is finite, then

$$|A| + |\bar{A}| = |S|. \tag{3}$$

Subsets and Power Sets

A set A is a *subset* of a set B (written $A \subseteq B$) if every element in A is also in B. Note then that $A \subseteq A$ for any set A and since one cannot prove that the empty set is not a subset of a set A, we must agree that $\emptyset \subseteq A$ for all sets A. In addition, if B is finite and $A \subseteq B$, then A must be finite with $|A| \leq |B|$. Given any set B, the set of all subsets of B is called the *power set* of B and is denoted by P_{B}. For example, if $B = \{1, 2, 3\}$, then

$$P_{\mathrm{B}} = \{\emptyset, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, B\}.$$

Note that if B is finite, then P_{B} is finite and

$$|P_{\mathrm{B}}| = 2^{|B|}. \tag{4}$$

In fact, if B is finite with $|B| = n$, then the number of subsets of B having exactly $m \leq n$ elements is the binomial coefficient

$$C(n, m) = \binom{n}{m} = \frac{n!}{m!(n-m)!} \quad \text{and of course} \quad \sum_{m=0}^{n} \binom{n}{m} = 2^n. \tag{5}$$

An example of this is summarized in the following table where $B = \{1, 2, 3, 4\}$.

Number of elements in each subset	Subsets of B	Number of subsets
0	$\emptyset$	$C(4, 0) = 1$
1	$\{1\}, \{2\}, \{3\}, \{4\}$	$C(4, 1) = 4$
2	$\{1, 2\}, \{1, 3\}, \{1, 4\}, \{2, 3\}, \{2, 4\}, \{3, 4\}$	$C(4, 2) = 6$
3	$\{1, 2, 3\}, \{1, 2, 4\}, \{1, 3, 4\}, \{2, 3, 4\}$	$C(4, 3) = 4$
4	$\{1, 2, 3, 4\}$	$C(4, 4) = 1$
	Totals	16

Equality of Sets

Two sets A and B are *equal* (written $A = B$) if A and B have the same elements. That is, if every element in A is also in B ($A \subseteq B$), and if every element in B is also in A ($B \subseteq A$). Thus we see that

$$A = B \qquad \textit{if and only if} \qquad A \subseteq B \qquad \text{and} \qquad B \subseteq A. \tag{6}$$

If $A \subseteq B$ and $A \neq B$, then A is called a *proper subset* of B and we write $A \subset B$. This says that all subsets of a set B are proper (including the empty set) except for the set B itself.

Intersection and Union of Sets

Given two sets A and B, the *intersection* of A and B is the set of all elements common to both A and B and is expressed as

$$A \cap B = \{x \,|\, x \in A \text{ and } x \in B\} \tag{7}$$

while the *union* of A and B is the set of all elements in both A or B and is expressed as

$$A \cup B = \{x \,|\, x \in A \text{ or } x \in B\}. \tag{8}$$

Note that all "or's" in this course are assumed to be *inclusive* (not exclusive). An *inclusive or* would be a statement like: "This semester I will take a course in mathematics *or* physics", because you can take both mathematics and physics. An *exclusive or* would be a statement like: "I will be in Philadelphia *or* in Boston at noon today", because you cannot be at both places at noon today. Two sets are called *mutually exclusive* or *disjoint* if they have no elements in common, *i.e.*, if $A \cap B = \emptyset$.

The Compliment of A in B or $B - A$

Given two sets A and B, the compliment of A in B is the set of all elements in B that are not in A and so this is the intersection of $\bar{A}$ and B, i.e.,

$$B \cap \bar{A} = \{x \,|x \in B \text{ and } \ x \notin A\} \tag{9}$$

and this is also written as $B - A$ so that $B \cap \bar{A} = B - A$.

2. A Review of Set Identities - The Algebra of Sets

Many *identities* involving sets can be derived from the above definitions, and many of these should be already familiar to the student. Since many identities between sets lead to corresponding identities involving probabilities, it is useful to record some of these set identities in this review.

Some Algebraic Identities

Some of the more useful algebraic identities include the following. In all of these, we assume that A, B and C are any three sets. We start with the obvious

$$A \cup A = A \qquad , \qquad A \cap A = A \tag{10}$$

and (*commutativity* of intersection and union)

$$A \cup B = B \cup A \qquad , \qquad A \cap B = B \cap A. \tag{11}$$

In addition we have (*associativity* of intersection and union)

$$(A \cup B) \cup C = A \cup (B \cup C) \quad \& \quad (A \cap B) \cap C = A \cap (B \cap C) \tag{12}$$

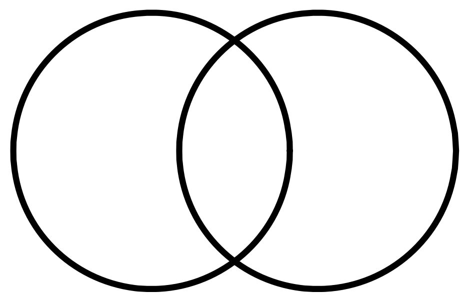
along with (*distributivity* between intersection and union)

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C) \tag{13a}$$

and

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C). \tag{13b}$$

Note that if A and B are finite, we may use the following figure as a guide,

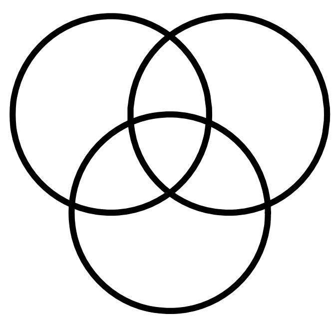


A is the set on the left (left disk), B is the set on the right (right disk) and S is the rectangle

to conclude that

$$|A \cup B| = |A| + |B| - |A \cap B| \tag{14a}$$

If A, B and C are all finite, we may use the following figure as a guide,



The Sets A (left disk), B (right disk), C (lower disk) and S (rectangle)

to conclude that

$$\begin{aligned} |A \cup B \cup C| &= |A| + |B| + |C| \\ &\quad -|A \cap B| - |A \cap C| - |B \cap C| \\ &\quad +|A \cap B \cap C|. \end{aligned} \tag{14b}$$

These two results are known as *inclusion-exclusion* relations and can be generalized to any number of sets, and this was already done in Chapter #2 of these notes. We conclude this subsection with the *DeMorgan's Rules* which state that

$$\overline{A \cup B} = \bar{A} \cap \bar{B} \qquad \text{and} \qquad \overline{A \cap B} = \bar{A} \cup \bar{B}. \tag{15a}$$

Note that DeMorgan's rules are true regardless of the choice in sample set S and these can be generalized to many sets to read

$$\overline{A_1 \cup A_2 \cup A_3 \cup \cdots \cup A_n} = \bar{A}_1 \cap \bar{A}_2 \cap \bar{A}_3 \cap \cdots \cap \bar{A}_n \tag{15b}$$

and

$$\overline{A_1 \cap A_2 \cap A_3 \cap \cdots \cap A_n} = \bar{A}_1 \cup \bar{A}_2 \cup \bar{A}_3 \cup \cdots \cup \bar{A}_n. \tag{15c}$$

Some Subset Identities

Once again, we assume that A, B and C are any three sets and that the symbol $\Leftrightarrow$ stands for *if and only if*, and we start with the elementary facts:

$$(A \cap B) \subseteq A \subseteq (A \cup B) \qquad \text{and} \qquad (A \cap B) \subseteq B \subseteq (A \cup B) \tag{16}$$

along with

$$A \subseteq B \Leftrightarrow (A \cup B) = B \Leftrightarrow (A \cap B) = A \tag{17a}$$

and

$$(A \subseteq C) \quad \text{and} \quad (B \subseteq C) \Leftrightarrow (A \cup B) \subseteq C \tag{17b}$$

and

$$(C \subseteq A) \quad \text{and} \quad (C \subseteq B) \Leftrightarrow C \subseteq (A \cap B). \tag{17c}$$

We also have the elementary identities:

$$(A \subseteq B) \Leftrightarrow (\bar{B} \subseteq \bar{A}) \tag{18a}$$

and

$$(A \cap B) = \emptyset \Leftrightarrow (A \subseteq \bar{B}) \Leftrightarrow (B \subseteq \bar{A}) \tag{18b}$$

and

$$(A \cup B) = S \Leftrightarrow (\bar{A} \subseteq B) \Leftrightarrow (\bar{B} \subseteq A). \tag{18c}$$

Partitions and Their Properties

Given a set A, a *partition* of A is a collection (or set) of *non-empty* subsets of A

$$\mathcal{P}(A) = \{A_1, A_2, A_3, \ldots, A_n\}$$

(n may be infinity) such that

$$A_i \cap A_j = \emptyset$$

for all $i \neq j$ and

$$A_1 \cup A_2 \cup A_3 \cup \cdots \cup A_n = A.$$

In particular, if $A \cap \bar{B}$, $A \cap B$ and $\bar{A} \cap B$ are all non-empty, then a natural partition of $A \cup B$ is given by

$$\mathcal{P}(A \cup B) = \{A \cap \bar{B}, A \cap B, \bar{A} \cap B\}$$

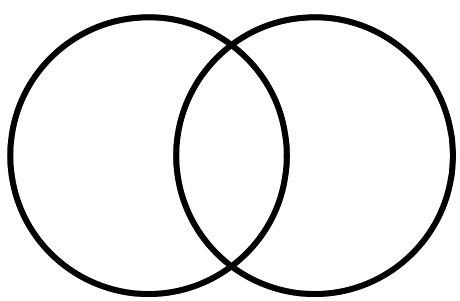
since

$$A \cup B = (A \cap \bar{B}) \cup (A \cap B) \cup (\bar{A} \cap B) \tag{19}$$

and

$$(A \cap \bar{B}) \cap (A \cap B) = (A \cap B) \cap (\bar{A} \cap B) = (A \cap \bar{B}) \cap (\bar{A} \cap B)$$

which are all equal to the empty set. These can be viewed in the following illustration.



$A \cap \bar{B}$ is on the left, $A \cap B$ is in the middle and $\bar{A} \cap B$ is on the right

The Cartesian Product

Given two sets A and B, the *Cartesian product* of A with B, denoted by $A \times B$ is given by

$$A \times B = \{(a,b) \mid a \in A \text{ and } b \in B\}. \tag{20a}$$

Note that in general, $A \times B \neq B \times A$ and if A and B are finite, then

$$|A \times B| = |A||B| = |B \times A|. \tag{20b}$$

They is much more we could say about sets but this is enough for now and so let us more on to some probability.

3. Some Probability Terminology and Definitions

Given an experiment, which is just the performance of something, the set of *all possible outcomes* of this experiment is called the *sample space* of the experiment and is denoted by S. The size of S, denoted by $|S|$ could be finite, infinite and countable, or infinite and uncountable. A possible outcome of the experiment being performed is called an event E and so an *event* E is simply a subset of S.

Example #1: Small Finite Sample Space

Suppose that an experiment consists of flipping a coin once, the sample space could be given by $S = \{H, T\}$, where H stands for heads and T stands for tails, and so $|S| = 2$. If the result of a flip is a heads, then $E = \{H\} \subseteq S$ would be a possible event. ■

Example #2: Two Different Finite Sample Spaces

Suppose that an experiment consists of flipping a coin twice (or two coins once), the sample space could be

$$S = \{(T,T), (T,H), (H,T), (H,H)\}$$

where H stands for heads and T stands for tails, if the result of the coin flips for both coins is recorded and here we have $|S| = 4$. If the result of the flips is a head follow by a tail, then

$$E = \{(H,T)\} \subseteq S$$

would be a possible event. One the other hand, suppose that only the total number of heads that appear is recorded, then the sample space would be

$$S = \{0, 1, 2\}$$

and now $|S| = 3$. If the result of the flips is a head follow by a tail, then, $E = \{1\} \subseteq S$ would be a possible event. ■

Example #3: Large Finite Sample Space

Suppose that the experiment is the order of finish in a race among 7 horses having positions 1, 2, 3, 4, 5, 6, 7, then

$$S = \{\text{all } 7! \text{ permutations of } (1, 2, 3, 4, 5, 6, 7)\}$$

so that $|S| = 7! = 5040$. If horse 3 wins the race, then

$$E = \{\text{all } 6! \text{ permutations of } (3, b, c, d, e, f, g)\} \subseteq S$$

would be a possible event. ■

Example #4: Infinite and Countable Sample Space

Suppose that an experiment consists of measuring the number (n) hits at a certain web site on a given day, then

$$S = \{0, 1, 2, 3, ...\}$$

showing that S is infinite in size and countable. If $120,000$ hits are made on a given day, then $E_1 = \{120,000\} \subseteq S$ would be a possible event. If less than $120,000$ hits were made on a given day, then $E_2 = \{0, 1, 2, 3, ..., 119,999\} \subseteq S$ would be another possible event. ■

Example #5: Infinite and Uncountable Sample Space

Suppose that an experiment consists of measuring (in hours) the lifetime (t) of a cell-phone battery, then

$$S = \{t | 0 \leq t < +\infty\}$$

showing that S is infinite in size and uncountable. If a battery is found to die after 20 hours of operation, then $E = \{t|0 \leq t \leq 20\} \subseteq S$ would be a possible event. ■

Example #6: Infinite and Uncountable Sample Space

Suppose that an experiment consists of measuring (in hours) the lifetime (t) of a cell-phone battery that is replaced after 25 hours, then

$$S = \{t|0 \leq t \leq 25 \text{ hours}\}$$

showing that S is infinite in size and uncountable. If a battery is found to still be working after 15 hours of operation, then $E = \{t|15 \leq t \leq 25\} \subseteq S$ would be a possible event. ■

The Limiting Relative Frequency Approach to Probability

One way of defining the probability of an event is in terms of its *relative frequency* of occurrence. Such a definition usually goes as follows. We suppose that an experiment, whose sample space is S, is repeatedly perform n times under exactly the same conditions, For each event E of the sample space S, we define $n(E)$ to be the number of times in the first n repetitions of the experiment that the event E occurs. Then $P(E)$, the probability of the event E, is defined as

$$P(E) = \lim_{n\to\infty} \frac{n(E)}{n}. \tag{21}$$

That is, $P(E)$ is defined as the (limiting) proportion of time that E occurs, which makes it the limiting relative frequency of E. Although this approach is certainly intuitively appealing, one should always keep in mind that it possesses a serious drawback, namely, how do we know that $n(E)/n$ will converge to some constant limiting value that will be the same for each possible sequence of repetitions of an experiment.

For example, suppose that the experiment to be repeatedly performed consists of flipping a coin. How do we know that the proportion of heads obtained in the first n flips will converge to some value as n gets large? Also, even if it converges to some value, how do we know that, if the experiment is repeatedly performed a second time (under the exact same conditions as the first), that we shall obtain the same limiting proportion of heads?

Proponents of the relative frequency definition of probability usually answer these objections by stating that the convergence of $n(E)/n$ to a constant limiting value is simply an assumption, or an *axiom*, of the system. However, to assume such an axiom is an extraordinarily complicated assumption. In addition, it is very hard to work with such an axiom when it comes to proving various theorems about probability. The more modern approach is to assume simpler and more self-evident axioms about probability and then to prove that such a constant limiting frequency exist under suitable conditions. In fact, we shall show that if $P(E)$ is the probability of an event E, and if $n(E)/n$ is the relative frequency on the number of occurrences of E in n measurements of the experiment, then, with probability 1, $n(E)/n$ approaches $P(E)$ in the limit as n gets large, *i.e.*,

$$P\left(\lim_{n\to\infty}\frac{n(E)}{n}=P(E)\right)=1. \tag{22}$$

This is the approach we shall take in this course but we should at least point out that the relative frequency approach is the basics behind using *simulation* in probability which is a very powerful method for estimation solutions to very complex probability problems that can not be easily solved for analytic solutions, and for checking analytic solutions to complex probability problems. We shall discuss this in much more detail later on.

The Axiomatic Approach - The Probability Function $P : \mathcal{P}_S \to [0,1]$

Suppose we have the sample space S of an experiment. We associate a function $P : \mathcal{P}_S \to [0,1]$ that maps each subset (or event) E of S (*i.e.*, elements of the power set of S) to a real number in the closed interval between 0 and 1, inclusive. We denote this number by $P(E)$ for $E \in \mathcal{P}_S$, and called it the *probability* that the event E occurs if an experiment is performed, and this probability function satisfies the following three (3) *axioms*:

(1) $0 \le P(E) \le 1$, for all events E in $\mathcal{P}_S$,

(2) $P(S) = 1$,

(3) If E_1, E_2, E_3, ..., E_n are pairwise mutually exclusive events, so that $E_i \cap E_j = \emptyset$ for $i \neq j$, then

$$P(E_1 \cup E_2 \cup \cdots \cup E_n) = P(E_1) + P(E_2) + \cdots + P(E_n). \tag{21a}$$

where n could be infinity (∞), which we may write as

$$P\left(\bigcup_{k=1}^{n} E_k\right) = \sum_{k=1}^{n} P(E_k) \qquad \text{or} \qquad P\left(\bigcup_{k=1}^{\infty} E_k\right) = \sum_{k=1}^{\infty} P(E_k).$$

Note that a special case of axiom (3) is for two events ($n = 2$) so that if E and F are mutually exclusive events, also called disjoint events (so that $E \cap F = \emptyset$), then

$$P(E \cup F) = P(E) + P(F). \tag{21b}$$

Axioms (1), (2) and (3) are known as the *probability axioms*, and these together with some basic properties of set theory lead to the results in the next section. We shall present these results along with their proofs.

Why are Probabilities Between Zero and One?

In many ways, the choice of $P : \mathcal{P}(S) \to [0, 1]$ is *completely arbitrary*. We could have chosen any finite closed interval and defined $P : \mathcal{P}(S) \to [a, b]$ (in which $a < b$) since the linear function $f : [a, b] \to [0, 1]$, where

$$f(x) = \frac{x - a}{b - a},$$

is a *one-to-one correspondence* between the two intervals $[a, b]$ and $[0, 1]$, which then says that the "size" of (or number of elements in) $[a, b]$ is the same as the "size" of (or number of elements in) $[0, 1]$. *The choice of the unit interval however, will allow for better agreement between the axiomatic description of probability and the relative frequency description.*

4. Some Probability Identities

The probability axioms can now be used to prove many basic properties about probability. Some of these are included in the following sets of results.

The Empty Set

If $\emptyset$ is the empty set, then $P(\emptyset) = 0$. To prove this, we simply note that $A \cup \emptyset = A$ for all sets A, and so

$$P(A \cup \emptyset) = P(A)$$

But A and $\emptyset$ are mutually exclusive (or disjoint) since $A \cap \emptyset = \emptyset$, and hence

$$P(A \cup \emptyset) = P(A) + P(\emptyset).$$

Thus we are lead to

$$P(A) + P(\emptyset) = P(A)$$

or $P(\emptyset) = 0$.

The Complementary Rule

If A is an event and if $\bar{A} = S - A$ (the complement of A in S), then

$$P(\bar{A}) = 1 - P(A). \tag{22}$$

To prove this, we simply note that for any subset A of S (*i.e.*, for any event A) $A \cup \bar{A} = S$ and so

$$P(A \cup \bar{A}) = P(S) = 1$$

But A and $\bar{A}$ (for any set A) are mutually exclusive (or disjoint) since $A \cap \bar{A} = \emptyset$, and hence

$$P(A \cup \bar{A}) = P(A) + P(\bar{A}).$$

Thus we are lead to

$$P(A) + P(\bar{A}) = 1$$

or $P(\bar{A}) = 1 - P(A)$.

The Union Rule: The Inclusion-Exclusion Principle

For any two events A and B, we have

$$P(A \cup B) = P(A) + P(B) - P(A \cap B). \tag{23a}$$

To prove this, we note that

$$(A \cap \bar{B}) \cup (A \cap B) \cup (\bar{A} \cap B) = A \cup B$$

is a partition of $A \cup B$ and so

$$P(A \cap \bar{B}) + P(A \cap B) + P(\bar{A} \cap B) = P(A \cup B).$$

But

$$(A \cap \bar{B}) \cup (A \cap B) = A \cap (\bar{B} \cup B) = A \cap S = A$$

is a partition of A and

$$(\bar{A} \cap B) \cup (A \cap B) = (\bar{A} \cup A) \cap B = S \cap B = B$$

is a partition of B. This says that

$$P(A \cap \bar{B}) + P(A \cap B) = P(A) \quad \text{and} \quad P(\bar{A} \cap B) + P(A \cap B) = P(B)$$

so that

$$P(A \cap \bar{B}) = P(A) - P(A \cap B) \qquad \text{and} \qquad P(\bar{A} \cap B) = P(B) - P(A \cap B)$$

and putting these into

$$P(A \cap \bar{B}) + P(A \cap B) + P(\bar{A} \cap B) = P(A \cup B).$$

we have

$$(P(A) - P(A \cap B)) + (P(B) - P(A \cap B)) + P(\bar{A} \cap B) = P(A \cup B)$$

resulting in Equation (23a). More generally, you should show that for any three events A, B and C,

$$\begin{aligned} P(A \cup B \cup C) &= P(A) + P(B) + P(C) \\ &\quad -P(A \cap B) - P(A \cap C) - P(B \cap C) \\ &\quad +P(A \cap B \cap C). \end{aligned} \tag{23b}$$

Equations (23a,b) are examples of what is known as the *inclusion-exclusion principle* and it should be clear that this can be generalized to any number of sets.

For example

$$\begin{aligned} P(A \cup B \cup C \cup D) &= P(A) + P(B) + P(C) + P(D) \\ &\quad -P(A \cap B) - P(A \cap C) - P(A \cap D) \\ &\quad -P(B \cap C) - P(B \cap D) - P(C \cap D) \\ &\quad +P(A \cap B \cap C) + P(A \cap B \cap D) \\ &\quad +P(A \cap C \cap D) + P(B \cap C \cap D) \\ &\quad -P(A \cap B \cap C \cap D) \end{aligned} \tag{23c}$$

and so on.

The Subset Rule

For any two events A and B, with $A \subseteq B$, we have $P(A) \leq P(B)$. To prove this, we start by noting that

$$A \cup (B \cap \bar{A}) = (A \cup B) \cap (A \cup \bar{A}) = (A \cup B) \cap S = A \cup B$$

and

$$A \cap (B \cap \bar{A}) = B \cap (A \cap \bar{A}) = B \cap \emptyset = \emptyset$$

so that

$$P(A \cup (B \cap \bar{A})) = P(A \cup B) = P(A) + P(B \cap \bar{A}).$$

But $A \cup B = B$ since we are given that $A \subseteq B$, and thus we have

$$P(B) = P(A) + P(B \cap \bar{A}) \geq P(A)$$

since $P(B \cap \bar{A}) \geq 0$ for any event $B \cap \bar{A}$. This result in intuitively appealing since it says that if B occurs, then A must also occur since $A \subseteq B$ and so B must be at least as probable as A.

Computing Probability Functions - Equally Likely Outcomes

Suppose that a sample space S can be *partitioned* into n outcomes: O_1, O_2, O_3, ..., O_n, so that

$$S = O_1 \cup O_2 \cup O_3 \cup \cdots \cup O_n$$

and $O_i \cap O_j = \varnothing$ for all $i \neq j$, and suppose that

$$P(O_i) = P(O_j)$$

for all $i, j, = 1, 2, 3, ..., n$, then we say that the outcomes, O_1, O_2, O_3, ..., O_n are *equally likely* and then

$$P(O_k) = \frac{1}{n} \tag{24a}$$

for each $k = 1, 2, 3, ..., n$. To prove this, we note that

$$S = O_1 \cup O_2 \cup O_3 \cup \cdots \cup O_n$$

and so

$$1 = P(S) = P(O_1 \cup O_2 \cup \cdots \cup O_n) = \sum_{k=1}^{n} P(O_k).$$

But since $P(O_i) = P(O_j)$, we must have $P(O_k) = \beta$ (a constant for each value of k). This leads to

$$1 = \sum_{k=1}^{n} P(O_k) = \sum_{k=1}^{n} \beta = n\beta$$

resulting in $\beta = 1/n$ and so $P(O_k) = \beta = 1/n$. In addition, if

$$E = \{O_{i_1}, O_{i_2}, O_{i_3}, ..., O_{i_k}\}$$

is an event consisting of k of these equally likely outcomes O_1, O_2, O_3, ..., O_n in S, then

$$P(E) = \frac{\text{number of equally likely outcomes in } E}{\text{number of equally likely outcomes in } S} = \frac{k}{n}. \tag{24b}$$

Example #7: Equally Likely Outcomes

Suppose that a *fair* coin is tossed once. There are two possible outcomes H (heads) and T (tails). If the sample space is $S = \{H, T\}$ and both heads and tails are equally likely to occur, then we must have $P(H) = P(T) = 1/2$. Of course if the coin is not fair so that $P(H) = p$ and $P(T) = 1 - p \neq p$, then the events H and T are not equally likely. ■

Example #8: Equally Likely Outcomes

Suppose that a *fair* six-sided die is tossed once. There are six possible outcomes numbered from 1 to 6. If the sample space is $S = \{1, 2, 3, 4, 5, 6\}$ and each number from 1 to 6 is equally likely to occur, then we must have

$$P(1) = P(2) = P(3) = P(4) = P(5) = P(6),$$

all equal to 1/6. Of course if the die is not fair, then the events 1, 2, 3, 4, 5, and 6 are not equally likely. ■

Example #9: Equally Likely Outcomes

Suppose that a *fair* coin is tossed twice and the result of each toss is recorded. There are four possible outcomes given by

$$S = \{(H,H),(H,T),(T,H),(T,T)\},$$

and these four events are equally likely to occur so that

$$P((H,H)) = P((H,T)) = P((T,H)) = P((T,T)) = \frac{1}{4}.$$

Of course if the coin is not fair so that $P(H) = p$ and $P(T) = 1 - p \neq p$, then these four events are not equally likely. ■

Example #10: Equally Likely Outcomes

Suppose that *two fair* six-sided die are tossed once and the result of each die is recorded. There are 36 possible outcomes,

$$\begin{aligned} S = \{&(1,1),(1,2),(1,3),(1,4),(1,5),(1,6),\\ &(2,1),(2,2),(2,3),(2,4),(2,5),(2,6),\\ &(3,1),(3,2),(3,3),(3,4),(3,5),(3,6),\\ &(4,1),(4,2),(4,3),(4,4),(4,5),(4,6),\\ &(5,1),(5,2),(5,3),(5,4),(5,5),(5,6),\\ &(6,1),(6,2),(6,3),(6,4),(6,5),(6,6)\}, \end{aligned}$$

and each of these 36 outcomes are equally likely to occur and so

$$P((i,j)) = \frac{1}{36},$$

for each $i,j = 1,2,3,4,5,6$. Of course if both or either of the two dice are not fair, then the 36 above are not equally likely. ■

Example #11: Not Equally Likely Outcomes

Suppose that a fair coin is tossed twice and the number of heads is recorded, Then there are three outcomes

$$S = \{0, 1, 2\},$$

but these three outcomes are not equally likely to occur since

$$P(0) = P((T,T)) = \frac{1}{4} \quad , \quad P(2) = P((H,H)) = \frac{1}{4}$$

and

$$P(1) = P((H,T) \cup (T,H)) = P((H,T)) + P((T,H)) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$$

and so we see that $P(0)$, $P(1)$ and $P(2)$ are not all equal. ■

Example #12: Not Equally Likely Outcomes

Suppose that two fair six-sided die are tossed once and the sum of each die is recorded. Then there are 11 possible outcomes

$$S = \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$$

and each of these 11 events are not equality likely to occur. The table below

E	Sets of Equally Likely Events	$P(E)$
2	$\{(1,1)\}$	1/36
3	$\{(1,2),(2,1)\}$	2/36
4	$\{(1,3),(2,2),(3,1)\}$	3/36
5	$\{(1,4),(2,3),(3,2),(4,1)\}$	4/36
6	$\{(1,5),(2,4),(3,3),(4,2),(5,1)\}$	5/36
7	$\{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\}$	6/36
8	$\{(2,6),(3,5),(4,4),(5,3),(6,2)\}$	5/36
9	$\{(3,6),(4,5),(5,4),(6,3)\}$	4/36
10	$\{(4,6),(5,5),(6,4)\}$	3/36
11	$\{(5,6),(6,5)\}$	2/36
12	$\{(6,6)\}$	1/36

shows that not all of $P(2)$, $P(3)$, ..., $P(12)$ are equal. ■

Example #13: Inclusion-Exclusion

Mary is taking two books along with her holiday vacation. With probability 0.5, she will like the first book, and with probability 0.4, she will like the second book, and with probability 0.3, she will like both books. Let us compute the probability that she likes neither book. To solve this, we let B_1 and B_2 be the events that she likes book 1 and book 2, respectively, and we are given that

$$P(B_1) = 0.5 \quad , \quad P(B_2) = 0.4 \quad \text{and} \quad P(B_1 \cap B_2) = 0.3,$$

and we want to compute

$$P(\bar{B}_1 \cap \bar{B}_2) = P(\overline{B_1 \cup B_2}) = 1 - P(B_1 \cup B_2).$$

But

$$P(B_1 \cup B_2) = P(B_1) + P(B_2) - P(B_1 \cap B_2) = 0.5 + 0.4 - 0.2 = 0.6$$

and so

$$P(\bar{B}_1 \cap \bar{B}_2) = 1 - P(B_1 \cup B_2) = 1 - 0.6 = 0.4$$

so that the probability that she likes neither book is 0.4. ■

Example #14: Using Combinatorics to Compute Probabilities

Suppose that 3 balls are *randomly drawn* (without replacement) from a bowl containing 6 white and 5 black balls. We want to compute the probability that one of the balls is white and the other two are black. If we regard the balls as being distinguishable so that the order in which they are selected is relevant, then (since there are $6 + 5 = 11$ distinguishable balls) the sample space consists of the

$$11 \times 10 \times 9 = 990$$

outcomes (a, b, c) where a, b and c are either white or black distinguishable balls. Furthermore, there are

$$WBB : 6 \times 5 \times 4 = 120$$

outcomes in which the first ball selected is white and the other two are black,

$$BWB : 5 \times 6 \times 4 = 120$$

outcomes in which the first ball selected is black, the second is white and the third is black, and

$$BBW : 5 \times 4 \times 6 = 120$$

outcomes in which the first two are black and the third is white. Hence, assuming that "randomly drawn" means that each outcome in the sample space is equally likely to occur, and since the events WBB, BWB and BBW are disjoint, we see that the desired probability is

$$P = \frac{120 + 120 + 120}{990} = \frac{4}{11}.$$

The problem could also have been solved by regarding the outcome of the experiment as the unordered set of drawn balls, making the balls indistinguishable, so that all 6 white balls look the same and all 5 black balls also look the same. From this point of view, there are now

$$\frac{11 \times 10 \times 9}{3!} = \binom{11}{3} = 165$$

outcomes in the sample space since now the events WBB, BWB and BBW are all the same. Now each set of 3 balls corresponds to $3! = 6$ outcomes when the order of selection is noted. As a result, if all outcomes are assumed equally likely when the order of selection is noted, then it follows that they remain equally likely when the outcome is taken to the unordered set of selected balls. Hence, using the latter representation of the experiment, there are

$$\binom{6}{1} = 6$$

ways in which the one white ball can be chosen from the 6, and there are

$$\binom{5}{2} = 10$$

ways in which the two black balls can be chosen from the 5, and so we see that the desired probability is

$$P = \frac{6 \times 10}{165} = \frac{4}{11},$$

which, of course, agrees with the answer obtained previously. ■

Equally Likely Events: Ordered Versus Unordered

When the experiment consists of a random selection of k items from a set of n items, we have the flexibility of either letting the outcome of the experiment be the ordered selection of k items or letting it be the unordered set of items selected. In the former case, we would assume that each new selection is equally likely to be any of the so far unselected items of the set, and in the latter case, we would assume that all $C(n, k)$ possible subsets of k items are equally likely to be the set selected.

Example #15: Ordered Versus Unordered

Suppose that 5 people are to be randomly selected from a group of 20 individuals consisting of 10 married couples, and we want to determine the probability that the 5 chosen are all unrelated (*i.e.*, no two are married to each other). If we regard the sample space as the set of 5 people chosen, then there are

$$\binom{20}{5} = 15,504$$

equally likely outcomes since there are 15,504 possible subsets of 5 elements that can be constructed from a set of 20 elements. An outcome that *does not* contain a married couple can be thought of as being the result of a *six-stage* experiment. In the first stage, look at the number of ways in which 5 couples

$$(C_{i_1}C_{i_2}C_{i_3}C_{i_4}C_{i_5})$$

can be chosen from a set of 10 couples

$$(C_1C_2C_3C_4C_5C_6C_7C_8C_9C_{10})$$

without replacement and without regard to order, and there are

$$\binom{10}{5} = 252$$

ways in which this can be done and it should be clear that people in different couples are certainly unrelated since only people in the same couple are related. In the next 5 stages, 1 of the 2 members of each of these couples

$$(C_{i_1}C_{i_2}C_{i_3}C_{i_4}C_{i_5})$$

is selected and there are $2 \times 2 \times 2 \times 2 \times 2 = 32$ ways in which this can be done. Thus, there are $252 \times 32 = 8,064$ possible outcomes in which the 5 members selected are unrelated, yielding the desired probability of

$$P = \frac{8,064}{15,504} = \frac{168}{323}.$$

In contrast, we could let the outcomes of the experiment be the ordered selection of the 5 individuals. In this setting, there are

$$20 \times (20-1) \times (20-2) \times (20-3) \times (20-4) = 1,860,480$$

equally likely outcomes of which

$$20 \times (20-2) \times (20-4) \times (20-6) \times (20-8) = 967,680$$

outcomes result in a group of 5 unrelated individuals, since each choice of a person must now eliminate 2 people from the remaining list (themselves and their spouse) if they are to be unrelated. This yields the probability

$$P = \frac{967,680}{1,860,480} = \frac{168}{323}$$

showing that the two answers are identical. ■

Example #16: Using Combinatorics

A committee of 5 is to be selected from a group of 6 men and 9 woman. If the selection is made randomly, let us compute the probability that the committee consist of 3 men and 2 woman. To do this we note that there are

$$\binom{6+9}{5} = \binom{15}{5} = 3003$$

different ways of forming a 5-person committee out of $6+9 = 15$ people. Note that a combination is used because the order of the 5-person committee is unimportant. There are

$$\binom{6}{3} = 20$$

different ways of choosing 3 men out of 6, and

$$\binom{9}{2} = 36$$

different ways of choosing 2 woman out of 9. Thus we find that

$$P = \frac{20 \times 36}{3003} = \frac{240}{1001}$$

o4 $P \simeq 0.24$ is the desired probability. ■

Example #17: Being Dealt a Straight in Poker

A poker hand consists of 5 cards from a deck of 52 cards with four suites (clubs, diamonds, hearts and spades) so that there are 13 clubs, 13 diamonds, 13 hearts and 13 spades and these range in value from 2, 3, 4, 5, 6, 7, 8, 9, 10, J, Q, K and A. If the cards have distinct consecutive values and are not all of the same suit, we say that the hand is a straight. For instance, a hand consisting of {5S, 6C, 7D, 8S, 9H} is a straight. Let us compute the probability of being dealt a straight. We begin be assuming that all the possible

$$\binom{52}{5} = 2,598,960$$

five-card hands in poker are equally likely to occur and to determine the number of outcomes that are straights, let us first determine the number of possible outcomes for which the poker hand consists of an A (ace), 2, 3, 4, 5 (the suites being irrelevant). Since the ace (A) can be any 1 of 4 possible aces (C, D, H, S), and similarly for the 2, 3, 4 and 5, it follows that there are

$$4 \times 4 \times 4 \times 4 \times 4 = 1,024$$

outcomes leading to exactly one A, 2, 3, 4 and 5. Also, since in 4 of these outcomes all the cards will be of the same suite (such a hand is called a straight flush), it follows that there are $4^5 - 4 = 1,020$ hands that make up a straight of the form {A, 2, 3, 4, 5} where the suits are different. Similarly, there are $1,020$ hands that make up a straight of the form {2, 3, 4, 5, 6}, {3, 4, 5, 6, 7}, and so one all the way up to {10, J, Q, K, A}, giving a total of $1,020 \times 10 = 10,200$ different possible hands that make up a straight (but not a straight flush). This we find that

$$P = \frac{10,200}{2,598,960} = \frac{5}{1274}$$

or $P = 0.0039 = 0.39\%$ gives the desired probability. ■

Probabilities in Poker hands

The student should verify the following probabilities when being dealt a five-card poker hand. (1) *A Single Pair*: This is the hand with the pattern AABCD, where A, B, C and D are from the distinct "kinds" of cards: aces, twos, threes, ..., tens, jacks, queens, and kings (there are 13 kinds, and four of each kind, in the standard 52 card deck). The number of such hands is

$$\binom{13}{1} \times \binom{4}{2} \times \binom{12}{3} \times \binom{4}{1} \times \binom{4}{1} \times \binom{4}{1} = 1,098,240$$

If all hands are equally likely, the probability of a single pair is obtained as

$$\frac{\binom{13}{1} \times \binom{4}{2} \times \binom{12}{3} \times \binom{4}{1} \times \binom{4}{1} \times \binom{4}{1}}{\binom{52}{5}} = \frac{352}{833} \simeq 0.42256903$$

(2) *Two Pairs*: This hand has the pattern AABBC where A, B, and C are from distinct kinds and the number of such hands is

$$\binom{13}{2} \times \binom{4}{2} \times \binom{4}{2} \times \binom{11}{1} \times \binom{4}{1} = 123,552.$$

If all hands are equally likely, the probability of getting two pairs is obtained as

$$\frac{\binom{13}{2} \times \binom{4}{2} \times \binom{4}{2} \times \binom{11}{1} \times \binom{4}{1}}{\binom{52}{5}} = \frac{198}{4165} \simeq 0.047539.$$

(3) *Three-of-a-Kind*: This hand has the pattern AAABC where A, B, and C are from distinct kinds, and the number of such hands is

$$\binom{13}{1} \times \binom{4}{3} \times \binom{12}{2} \times \binom{4}{1} \times \binom{4}{1} = 54,912.$$

If all hands are equally likely, the probability of getting three-of-a-kind is obtained as

$$\frac{\binom{13}{1} \times \binom{4}{3} \times \binom{12}{2} \times \binom{4}{1} \times \binom{4}{1}}{\binom{52}{5}} = \frac{88}{4165} \simeq 0.021128.$$

(4) *A Full House*: This hand has the pattern AAABB where A and B are from distinct kinds, and the number of such hands is

$$\binom{13}{1} \times \binom{4}{3} \times \binom{12}{1} \times \binom{4}{2} = 3,744.$$

If all hands are equally likely, the probability of getting a full house is obtained as

$$\frac{\binom{13}{1} \times \binom{4}{3} \times \binom{12}{1} \times \binom{4}{2}}{\binom{52}{5}} = \frac{6}{4165} \simeq 0.0014405762.$$

(5) *Four-of-a-Kind*: This hand has the pattern AAAAB where A and B are from distinct kinds, and the number of such hands is

$$\binom{13}{1} \times \binom{4}{4} \times \binom{12}{1} \times \binom{4}{1} = 624.$$

If all hands are equally likely, the probability of getting four-of-a-kind is obtained as

$$\frac{\binom{13}{1} \times \binom{4}{4} \times \binom{12}{1} \times \binom{4}{1}}{\binom{52}{5}} = \frac{1}{4165} \simeq 0.0002401.$$

(6) *A Straight*: This is five cards in a sequence such as 4,5,6,7,8, with aces allowed to be either 1 or 13 (low or high) and with the cards allowed to be of the same suit (*e.g.*, all hearts) or from some different suits. The number of such hands is

$$\binom{4}{1} \times \binom{4}{1} \times \binom{4}{1} \times \binom{4}{1} \times \binom{4}{1} \times 10 = 10,240.$$

If all hands are equally likely, the probability of getting a straight is obtained as

$$\frac{\binom{4}{1} \times \binom{4}{1} \times \binom{4}{1} \times \binom{4}{1} \times \binom{4}{1} \times 10}{\binom{52}{5}} = \frac{128}{32487} \simeq 0.00394.$$

In this calculation, if one means to exclude straight flushes and royal flushes, then number of such hands is

$$\binom{4}{1} \times \binom{4}{1} \times \binom{4}{1} \times \binom{4}{1} \times \binom{4}{1} \times 10 - 36 - 4 = 10,200.$$

If all hands are equally likely, the probability of getting only a straight is obtained as

$$\frac{\binom{4}{1} \times \binom{4}{1} \times \binom{4}{1} \times \binom{4}{1} \times \binom{4}{1} \times 10 - 36 - 4}{\binom{52}{5}} = \frac{5}{1274} \simeq 0.00392465.$$

(7) *A Flush*: Here all 5 cards are from the same suit and they may also be a straight. The number of such hands is

$$\binom{4}{1} \times \binom{13}{5} = 5,148.$$

If all hands are equally likely, the probability of getting a flush is obtained as

$$\frac{\binom{4}{1} \times \binom{13}{5}}{\binom{52}{5}} = \frac{33}{16660} \simeq 0.00198079.$$

In this calculation, if one means to exclude straight flushes, then the number of such hands is

$$\binom{4}{1} \times \binom{13}{5} - 4 \times 10 = 5,108$$

and if all hands are equally likely, the probability of getting only a flush is obtained as

$$\frac{\binom{4}{1} \times \binom{13}{5} - 4 \times 10}{\binom{52}{5}} = \frac{1277}{649740} \simeq 0.0019654.$$

(8) *A Straight Flush*: Here, all 5 cards are from the same suit and they form a straight (they may also be a royal flush). The number of such hands is $4 \times 10 = 40$, and if all hands are equally likely, the probability of getting a straight flush is obtained as

$$\frac{4 \times 10}{\binom{52}{5}} = \frac{1}{64974} \simeq 0.00001539.$$

In this calculation, if one means to exclude royal flushes, then the number of such hands is $4 \times 10 - 4 = 36$, and if all hands are equally likely, the probability of getting only a straight flush is obtained as

$$\frac{4 \times 10 - 4}{\binom{52}{5}} = \frac{3}{216580} \simeq 0.0000138516\,95.$$

(9) *A Royal Flush*: This consists of the ten, jack, queen, king, and ace of one suit. There are four such hands and if all hands are equally likely, the probability of getting a royal flush is

$$\frac{4}{\binom{52}{5}} = \frac{1}{649740} \simeq 0.000001539.$$

(10) *None of the Above*: We have to choose 5 distinct kinds

$$\binom{13}{5}$$

but exclude any straights

$$\binom{13}{5} - 10.$$

We can have any pattern of suits except the 4 patterns where all 5 cards have the same suit ($4^5 - 4 = 1,020$). The total number of such hands is

$$\left(\binom{13}{5} - 10\right) \times (4^5 - 4) = 1,302,540,$$

and if all hands are equally likely, the probability of getting nothing is obtained as

$$\frac{\left(\binom{13}{5} - 10\right) \times (4^5 - 4)}{\binom{52}{5}} = \frac{1277}{2548} \simeq 0.50118.$$

These are all summarized in the following table of probabilities. ***

Hand	Probability
Nothing	0.510118
Single Pair	0.422569
Two Pairs	0.047539
Three-of-a-Kind	0.021128
Straight	0.003940
Flush	0.001965
Full House	0.001441
Four-of-a-Kind	0.000240
Straight Flush	0.000014
Royal Flush	0.000002

Of course, we may check these by summing over all the probabilities involving disjoint sets to get

$$\frac{352}{833} + \frac{198}{4165} + \frac{88}{4165} + \frac{6}{4165} + \frac{1}{4165} + \frac{5}{1274}$$
$$+ \frac{1277}{649740} + \frac{3}{216580} + \frac{1}{649740} + \frac{1277}{2548}$$

which gives 1. One sees that winning in poker would not be so frequent if it were not for bluffing! ■

Example #18: A Distribution of Runs

Consider an athletic team that had just finished its season with a final record of n wins and m losses. By examining the sequence of wins and losses, we are hoping to determine whether the team had stretches of games in which it was more likely to win at one time than at some other time. One way to gain some insight into this question is to count the number of runs of wins and then see how likely that result would be when all $(n+m)!/(n!m!)$ orderings of n wins and m losses are assumed equally likely. By a run of wins, we mean a consecutive sequence of 1 or more wins. For instance, suppose that $n = 10$ and $m = 6$ and the sequence of $10 + 6 = 16$ outcomes was

$$WWLLWWWLWLLLWWWW = [WW]LL[WWW]L[W]LLL[WWWW]$$

then there would be 4 runs of wins (shown in square brackets). The first run $[WW]$, is said to be of length 2, the second $[WWW]$, of length 3, the third $[W]$, of length 1 and the fourth $[WWWW]$, of length 4.

Suppose now that a team had n wins and m losses. Assuming that all

$$\binom{n+m}{n} = \frac{(n+m)!}{n!m!}$$

orderings of n wins and m losses are equally likely, let us determine the probability that there will be exactly r runs of wins. First we note that the smallest possible value of r is $r = 1$, which occurs when all runs are of length n, which is when all n wins occur all at once at the beginning followed by all the m losses at the end, such as in

$$WW\cdots WWLL\cdots LL = [WW\cdots WW]LL\cdots LL$$

or when all n wins occur at the end proceeded by all m losses in the beginning, such as in

$$LL\cdots LLWW\cdots WW = LL\cdots LL[WW\cdots WW]$$

which are both single runs $(r = 1)$ of length n. The largest possible value of r is

$$r = \min(n, m+1)$$

which occurs when all runs of wins are of length 1 such as in

$$WLWLLWLL = [W]L[W]LL[W]LL$$

when $n = 3$ and $m = 5$, or when all runs of losses are of length 1 such as in

$$WLWLWLWW = W[L]W[L]W[L]WW$$

when $n = 5$ and $m = 3$.

To determine the above probability, we must do some counting and to do this, let us determine a *one-to-one correspondence* to this number, which is easier to count. Toward this end, consider the example above in which

$$WWLLWWWLWLLLWWWW = [WW]LL[WWW]L[W]LLL[WWWW].$$

We may associate with this sequence of W's and L's a set of integers

$$(x_1, x_2, x_3, x_4) = (2, 3, 1, 4) \quad \text{and} \quad (y_1, y_2, y_3, y_4, y_5) = (0, 2, 1, 3, 0)$$

or we may associate with the sequence of W's and L's

$$LWWLWWLWWLWWWWLL = L[WW]L[WW]L[WW]L[WWWW]LL,$$

the set of integers

$$(x_1, x_2, x_3, x_4) = (2, 2, 2, 4) \quad \text{and} \quad (y_1, y_2, y_3, y_4, y_5) = (1, 1, 1, 1, 2).$$

Note that

$$x_1 + x_2 + x_3 + x_4 = 10 = n \quad \text{and} \quad y_1 + y_2 + y_3 + y_4 + y_5 = 6 = m$$

is both of these examples.

In general, we may associated with the sequence of W's and L's

$LL...L$ y_1 of these	$WW...W$ x_1 of these	$LL...L$ y_2 of these	$WW...W$ x_2 of these	...	$WW...W$ x_r of these	$LL...L$ y_{r+1} of these

a vector of *positive* integers

$$(x_1, x_2, x_3, ..., x_r) \quad \text{with} \quad x_1 + x_2 + x_3 + \cdots + x_r = n$$

which satisfy $x_i > 0$ for $i = 1, 2, 3, 4, ..., r$, and a vector of *non-negative* integers

$$(y_1, y_2, y_3, ..., y_{r+1}) \quad \text{with} \quad y_1 + y_2 + y_3 + \cdots + y_{r+1} = m$$

which satisfy: $y_1 \geq 0$, $y_{r+1} \geq 0$ and $y_j > 0$ (or $y_j \geq 1$) for $j = 2, 3, 4, ..., r$. Note that each x_i gives the length of the ith run of W's and each y_j gives the length of the jth run of L's with the possibility that $y_1 = 0$ when the first outcome is a win and $y_{r+1} = 0$ when the last outcome is a win. For any such outcome of wins we see that y_1 denotes the number of losses before the first run of wins, y_2 the number of losses between the first 2 runs of wins, ..., y_{r+1} the number of losses after the last run of wins.

Let us now see how many outcomes result in r runs of wins in which the ith run of W's is of size x_i, with $i = 1, 2, 3, ..., r$ and the jth run of L's is of size y_j, with $j = 1, 2, 3, ..., r + 1.$, with $y_1 = 0$ when the first outcome is a win and $y_{r+1} = 0$ when the last outcome is a win and $y_j > 0$ for all $j = 2, 3, ..., r$. Using Equation (7c) from Chapter #2, we know that if c_1, c_2, c_3, ..., c_n are integers (positive, negative or zero), then the number of different solutions (in integers) to the equation

$$z_1 + z_2 + z_3 + \cdots + z_k = n$$

with the conditions that $x_i > c_i$ for all $i = 1, 2, 3, ..., k$ is

$$\binom{n - c_1 - c_2 - c_3 - \cdots - c_n - 1}{k - 1}.$$

Using this result, the total number of *positive* integer solutions to

$$x_1 + x_2 + x_3 + \cdots + x_r = n \qquad \text{is} \qquad \binom{n-1}{r-1} \tag{25a}$$

since in this case $c_i = 0$ for $i = 1, 2, 3, ..., r$. Using the same result, we see that the total number of integer solutions to

$$y_1 + y_2 + y_3 + \cdots + y_{r+1} = m$$

which satisfy $y_1 \geq 0 > -1 = c_1$, $y_{r+1} \geq 0 > -1 = c_{r+1}$ and $y_j > 0 = c_j$ for $j = 2, 3, 4, ..., r$, is

$$\binom{m - (-1) - (-1) - 1}{(r+1) - 1} = \binom{m+1}{r}. \tag{25b}$$

Therefore, the total number of possible sequences of n (W's) and m (L's) is then

$$\binom{n-1}{r-1} \times \binom{m+1}{r}$$

and since each of these is equally likely to occur, we find that

$$P(r \text{ runs of wins}) = \frac{\binom{m+1}{r}\binom{n-1}{r-1}}{\binom{n+m}{n}} \tag{26}$$

for $r = 1, 2, 3, ..., \min(n, m+1)$ gives the desired probability. ■ As an application of Equation (26), we may look at the next example.

Example #19: The Systems of Antennas Problem - An Application

A communication system is to consist of N seemingly identical antennas that are lined up in a linear order. The resulting system will then be able to receive all incoming signals and will be called *functional* as long as no two consecutive antennas are defective. If it turns out that exactly M of the N antennas are defective (with $M \leq N$), we would like to compute the probability that the resulting system will be functional. Let us represent a defective antenna by a 0 and a working antenna by a 1. First we note that if there are M (0's) and hence $N - M$ (1's), we must have

$$N - M \geq M - 1 \qquad \text{or} \qquad N \geq 2M - 1,$$

or else there is no way to avoid having at least two zeros in a row and hence no way to avoid having at least two consecutive antennas that are defective. For example, if $N = 8$ and $M = 5$, then $N - M = 3$ and placing down the 5 zeros (00000), there is no way to place down the 3 ones (111) between these 5 zeros without having at least two zeroes in a row. We would need at least 4 ones (1111) for then we may place them between the 5 zeros and get (010101010). Thus we may state right from the start that

$$P(\text{system is functional}) = 0 \qquad \text{when} \qquad N < 2M - 1.$$

To see how to solve this problem when $N \geq 2M - 1$, let us now look at a few examples. For instance, in the special case where $N = 4$ and $M = 2$, the possible system configurations are

0011	1001
0101	1010
0110	1100

where 1 means the antenna is working and 0 means the antenna is defective. Of these 6 possibilities,

0011	1001	1100

shows at least two defective antennas in a row, and hence shows the cases where the linear array of 4 antennas is not functional, while

0110	0101	1010

show no two defective entries in a row and hence shows the cases where the linear array of 4 antennas is functional. This says that

$$P(\text{system is functional}) = \frac{3}{6} = \frac{1}{2}.$$

As another example, suppose that $N = 7$ and $M = 3$, then the different possible system configurations are shown in the table below.

0001111	0101101	0111100	1010110	1101001
0010111	0101110	1000111	1011001	1101010
0011011	0110011	1001011	1011010	1101100
0011101	0110101	1001101	1011100	1110001
0011110	0110110	1001110	1100011	1110010
0100111	0111001	1010011	1100101	1110100
0101011	0111010	1010101	1100110	1111000

The ones having no two 0's in a row are

0101011	0111010
0101101	1010101
0101110	1010110
0110101	1011010
0110110	1101010

and this says that

$$P(\text{system is functional}) = \frac{10}{35} = \frac{2}{7}$$

It should be clear that the system is functional as long as all runs of 0's are of length 1 because any run of length 2 or more will result in a non-functional system of antennas. Using the result of the previous example with W's replaced by 0's and L's replaced by 1's, we have

$$n = \text{number of } W\text{'s} = \text{number of 0's} = M$$

and

$$m = \text{number of } L\text{'s} = \text{number of 1's} = N - M.$$

Using Equation (26), we have

$$P(r \text{ runs of wins}) = \frac{\binom{m+1}{r}\binom{n-1}{r-1}}{\binom{n+m}{n}}$$

and we want all runs to be of length 1 since a runs of 0's of length 2 or more results in a non-functional system of antennas. This says that

$$r = \min(n, m+1) = \min(M, N - M + 1),$$

and since we are assuming that $N \geq 2M - 1$, we have $N - M + 1 \geq M$ and so

$$r = \min(M, N - M + 1) = M$$

This we find that

$$P(\text{system is functional}) = \frac{\binom{N-M+1}{r}\binom{M-1}{r-1}}{\binom{N}{M}} = \frac{\binom{N-M+1}{M}\binom{M-1}{M-1}}{\binom{N}{M}}$$

which reduces to

$$P(\text{system is functional}) = \frac{(N-M+1)!M!(N-M)!}{M!(N+1-2M)!N!}$$

or

$$P(\text{system is functional}) = \frac{(N-M+1)!(N-M)!}{N!(N+1-2M)!} \quad \text{when} \quad N \geq 2M - 1.$$

Thus we find that $P(\text{system is functional}) = 0$ when $N < 2M - 1$, and

$$P(\text{system is functional}) = \frac{(N-M+1)!(N-M)!}{N!(N+1-2M)!} \quad \text{when} \quad N \geq 2M - 1.$$

As a check, we note that using $N = 4$ and $M = 2$ gives

$$P(\text{system is functional}) = \frac{(4-2+1)!(4-2)!}{4!(4+1-2(2))!} = \frac{3!2!}{4!1!} = \frac{1}{2}$$

while using $N = 7$ and $M = 3$, we have

$$P(\text{system is functional}) = \frac{(7-3+1)!(7-3)!}{7!(7+1-2(3))!} = \frac{5!4!}{7!2!} = \frac{2}{7}$$

and these both agree with the results above. ■

The Factorial Function

Later in this course, it will become convenient to use the factorial function defined as

$$x! = \int_0^\infty t^x e^{-t} dt \tag{27}$$

for $x \geq 0$. To understand the motivation for this definition, we note that (via integration by parts)

$$x! = \int_0^\infty t^x e^{-t} dt = -t^x e^{-t}|_0^\infty + \int_0^\infty x t^{x-1} e^{-t} dt$$

and since

$$\lim_{t\to\infty} (t^x e^{-t}) = 0 \qquad \text{and} \qquad \lim_{t\to 0^+} (t^x e^{-t}) = 0$$

for $x \geq 0$, we have

$$x! = \int_0^\infty x t^{x-1} e^{-t} dt = x \int_0^\infty t^{x-1} e^{-t} dt$$

or

$$x! = x(x-1)! \tag{28}$$

for $x \geq 0$. For $x = 0$, we have

$$0! = \int_0^\infty t^0 e^{-t} dt = \int_0^\infty e^{-t} dt = -e^{-t}|_0^\infty = 1$$

and so, using Equation (28), we have

$$1! = 1(0!) = 1 \quad , \quad 2! = 2(1!) = 2 \quad , \quad 3! = 3(2!) = (3)(2)(1) = 6$$

and it is easy to see that

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

for $n = 1, 2, 3, ...,$ as we had defined earlier. What Equation (27) does is to extend the definition of factorial beyond just integers. For example

$$(1/2)! = \int_0^\infty t^{1/2}e^{-t}dt = \frac{1}{2}\sqrt{\pi} \simeq 0.8862.$$

In fact, it can be shown that Equation (27) is well defined and continuous for all x expect when $x = -1, -2, -3, ...,$ and $x! = \pm\infty$ (or $1/x! = 0$) for $x = -1, -2, -3, ...$

Using this factorial function, we may now take the result in Example #19 and say that

$$P(\text{system is functional}) = \frac{(N-M+1)!(N-M)!}{N!(N+1-2M)!}$$

for all N and M, since the

$$\frac{1}{(N+1-2M)!}$$

will automatically give zero when $N < 2M - 1$. Let us now look at two more examples.

Example #20: A Surprise

Suppose that you have a litter of four cats. Is the litter having equal number of males (two) and females (two) the most probable? To answer this, let us suppose that S is the sample space consisting of the number of cats that have the same sex. Thus we find that

$$S = \{2, 3, 4\}$$

where

$$P(2) = P((FFMM)) = \binom{4}{2}\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^2 = \frac{3}{8}$$

and

$$P(3) = P((FFFM)) + P((FMMM)) = \binom{4}{3}\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)^1 + \binom{4}{3}\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)^1$$

or $P(3) = 4/8 = 1/2$, and

$$P(4) = P((FFFF)) + P((MMMM)) = \binom{4}{4}\left(\frac{1}{2}\right)^4 + \binom{4}{4}\left(\frac{1}{2}\right)^4 = \frac{1}{8}$$

This shows that having three of the cats the same sex is the most probable. ■

Example #21: Asking An Embarrassing Question

Quite often people in the medical profession must ask embarrassing questions to a group of people, and as Dr. House says on the TV drama *House* "Everybody lies". So how can we get people in a group to truthfully answer an embarrassing question. One way is to use probability and statistics in the following way. Suppose we have a group of N people and suppose that E of these N people perform some embarrassing act that they do not want to admit in a public forum. Our goal is to determine an estimate for E. We can easily determine N since we simply count the number of people in the forum. It would be nice to simply ask those people that perform the embarrassing act to raise their hand and count the number of hands up to get E. But we know that many will not want to admit in public (or in private) that they perform the embarrassing act. Instead, we ask each of the N people to take a fair coin and flip it without showing the result of the flip to anyone but themselves. Then we ask them to raise their hand if they (1) either perform the embarrassing act, or (2) get a heads on their coin flip. This way a person having their hand raised does not reveal that they perform the embarrassing act since their hand could be raised because their coin flip was heads (which only they can see). This allows the people to now be honest. To see how this can be used to estimate the value of E, suppose there are H (of the N) hands up in the air, then a person's hand will be up in the air if they perform the embarrassing act or if they don't and got a heads on their coin flip. Therefore, the average number of hands that should be up are

$$E + \frac{1}{2}(N - E) = H = \frac{1}{2}(N + E).$$

Solving this for E, we have

$$E = 2H - N$$

as an estimate for E. However, this is not a very good estimate for E because it represents a *single sample estimate* for E, since it is based on just a single flip of each person's coin. To get a getter estimate, we could ask the people to do the above process many times, (say n times) and suppose that H_i equals the number of hands raised during the ith try in which $i = 1, 2, 3, ..., n$. Then $E_i = 2H_i - N$ for $i = 1, 2, 3, ..., n$, and

$$E = \frac{1}{n}\sum_{i=1}^{n} E_i = \frac{1}{n}\sum_{i=1}^{n}(2H_i - N) = 2\left(\frac{1}{n}\sum_{i=1}^{n} H_i\right) - N = 2H_{\text{ave}} - N$$

or

$$E \simeq 2H_{\text{ave}} - N \qquad \text{where} \qquad H_{\text{ave}} = \frac{1}{n}\sum_{i=1}^{n} H_i$$

equals the average number of hands raised in n tosses of the coins. For example, suppose in a room of $N = 100$ people, we have them toss the coin $n = 10$ times and get the following results.

i	1	2	3	4	5	6	7	8	9	10
H_i	65	70	63	71	68	61	66	73	59	68

Then

$$H_{\text{ave}} = \frac{65 + 70 + 63 + 71 + 68 + 61 + 66 + 73 + 59 + 68}{10} = 66.4$$

and so $E \simeq 2(66.4) - 100 = 32.8$ (or $E \simeq 33$) of the 100 people perform the embarrassing act. Of course, doing something like this does raise a problem since if one person is watching another person and sees that person's hand is raised for all $n = 10$ trials, then there is a very good chance that the person whose hand is always raised performs the embarrassing act. On the other hand (no pun intended), if another person's hand is raised during one trial and not during another, then it is certain that this person does not perform the embarrassing act. Therefore, having the people move around the room during each coin toss is probably not a bad idea, or using today's technology, having them transmit their "hand raises" electronically is even a better idea. In fact, using today's technology, we can even have the people provide honest answers without knowing who supplied the answer. ■

Example #22: The Birthday Problem - A Real Surprise

Suppose that there are n people in a room. What is the probability that at least two of the people have the same birthday, assuming a 365-day year?

To answer this, let $N = 365$ be the number of days in a year and let E be the event that at least two people have the same birthday and consider $\bar{E}$ which is the event that no two people have the same birthday. Then there are N days to choose for the first person leaving $N-1$ days for the second, leaving $N-2$ days for the third, and so on until we have $N-(n-1)$ for the nth person. This leads to a probability of

$$P(\bar{E}) = \frac{(N)(N-1)(N-2)\cdots(N-(n-1))}{(N)(N)(N)\cdots(N)} = \frac{N!}{N^n(N-n)!}.$$

and this says that

$$P(E) = 1 - P(\bar{E}) = 1 - \frac{N!}{N^n(N-n)!}.$$

For $N = 365$, we have

$$P(E) = 1 - \frac{(365)!}{(365)^n(365-n)!}$$

and a short table of these values in shown as follows.

n	$P(E)$	n	$P(E)$
5	0.0271	50	0.9704
10	0.1169	55	0.9863
15	0. 2529	60	0.9941
20	0.4114	65	0.9977
22	0.4757	70	0.9992
23	**0.5073**	75	0.9997
25	0.5687	80	0.9999
30	0.7063	85	0.999976
35	0.8144	90	0.999994
40	0.8912	95	0.9999986
45	0.9410	100	0.9999997

The surprising this about this result is that the probability is larger than 1/2 when there are only 23 people in the room! Many people have been known to make money at a party by betting all the people there that at least two of the people at the party have the same birthday. For a party having only 40 people, your probability of winning is just shy under 90% and so your chances of winning are quite large. ■

Probability as a Continuous Set Function - Optional

A sequence of events $\{I_n$ for $n = 1, 2, 3, ...\}$ is said to be an *increasing* sequence if

$$I_1 \subset I_2 \subset I_3 \subset \cdots \subset I_n \subset I_{n+1} \subset \cdots$$

and note that in this case

$$I_1 \cup I_2 \cup I_3 \cup \cdots \cup I_n = I_n \quad \text{and} \quad I_1 \cap I_2 \cap I_3 \cap \cdots \cap I_n = I_1$$

for each $n = 1, 2, 3, ...$ A sequence of events $\{D_n$ for $n = 1, 2, 3, ...\}$ is said to be *decreasing* if

$$D_1 \supset D_2 \supset D_3 \supset \cdots \supset D_n \supset D_{n+1} \supset \cdots$$

and note that in this case

$$D_1 \cup D_2 \cup D_3 \cup \cdots \cup D_n = D_1 \quad \text{and} \quad D_1 \cap D_2 \cap D_3 \cap \cdots \cap D_n = D_n$$

for each $n = 1, 2, 3, ...$

If a sequence of events $\{I_n$ for $n = 1, 2, 3, ...\}$ is an increasing sequence of events, then we define the new event

$$I_\infty = \lim_{n\to\infty} I_n = \bigcup_{k=1}^{\infty} I_k \tag{29a}$$

(assuming that the limit does exists) and if a sequence of events $\{D_n$ for $n = 1, 2, 3, ...\}$ is a decreasing sequence of events, then we define the new event

$$D_\infty = \lim_{n\to\infty} D_n = \bigcap_{k=1}^{\infty} D_k, \tag{29b}$$

(assuming again that the limit does exists). Using these definitions for I_n and D_n (for an increasing or decreasing sequence of events), let us prove that

$$E_\infty = \lim_{n\to\infty} E_n \quad \text{implies that} \quad P(E_\infty) = \lim_{n\to\infty} P(E_n) \tag{30}$$

for $E_n = I_n$ and for $E_n = D_n$.

Toward this end, suppose that $\{E_n \text{ for } n = 1, 2, 3, ...\}$ is an *increasing* sequence of events, and define the events F_n for $n = 1, 2, 3, ...$, by $F_1 = E_1$, and

$$F_n = E_n \cap (\overline{E_1 \cup E_2 \cup E_3 \cup \cdots \cup E_{n-1}}) = E_n \cap \overline{E}_{n-1}$$

where we have used the fact that

$$E_1 \cup E_2 \cup E_3 \cup \cdots \cup E_{n-1} = E_{n-1}$$

for an increasing sequence of events. In words, F_n (for $n = 2, 3, 4, ..,$) consists of those outcomes of E_n that are not in any of the earlier E_i's for $i = 1, 2, 3, ..., n-1$. It is easy to see that

$$F_1 \cup F_2 = E_1 \cup (E_2 \cap \overline{E}_1) = (E_1 \cup E_2) \cap (E_1 \cup \overline{E}_1) = (E_1 \cup E_2) \cap S$$

or

$$F_1 \cup F_2 = E_1 \cup E_2$$

and then

$$\begin{aligned}
F_1 \cup F_2 \cup F_3 &= (E_1 \cup E_2) \cup F_3 = (E_1 \cup E_2) \cup (E_3 \cap \overline{E}_2) \\
&= (E_1 \cup E_2 \cup E_3) \cap (E_1 \cup E_2 \cup \overline{E}_2) \\
&= (E_1 \cup E_2 \cup E_3) \cap (E_1 \cup S) \\
&= (E_1 \cup E_2 \cup E_3) \cap S = E_1 \cup E_2 \cup E_3
\end{aligned}$$

and so on, so that

$$F_1 \cup F_2 \cup F_3 \cup \cdots \cup F_n = E_1 \cup E_2 \cup E_3 \cup \cdots \cup E_n$$

for all $n = 1, 2, 3, ...$, and in the limit we have

$$\bigcup_{k=1}^{\infty} F_k = \bigcup_{k=1}^{\infty} E_k.$$

In addition, we note that if $i < j$, then

$$
\begin{aligned}
F_i \cap F_j &= (E_i \cap \overline{E}_{i-1}) \cap (E_j \cap \overline{E}_{j-1}) \\
&= (E_i \cap E_j) \cap (\overline{E}_{i-1} \cap \overline{E}_{j-1}) \\
&= E_i \cap (\overline{E_{i-1} \cup E_{j-1}}) \\
&= E_i \cap \overline{E_{j-1}} = \emptyset
\end{aligned}
$$

because E_{j-1} (for $i < j$ and hence $i \leq j-1$) contains all elements of E_i and hence $\overline{E_{j-1}}$ contains none of the elements in E_i. This says that all the F_i's are mutually exclusive events and so

$$
P\left(\bigcup_{k=1}^{\infty} E_k\right) = P\left(\bigcup_{k=1}^{\infty} F_k\right) = \sum_{k=1}^{\infty} P(F_k)
$$

or

$$
P(E_\infty) = \sum_{k=1}^{\infty} P(F_k) = \lim_{n\to\infty} \sum_{k=1}^{n} P(F_k) = \lim_{n\to\infty} P\left(\bigcup_{k=1}^{n} F_k\right) = \lim_{n\to\infty} P\left(\bigcup_{k=1}^{n} E_k\right)
$$

or

$$
P(E_\infty) = \lim_{n\to\infty} P(E_n)
$$

and the proof is complete. A similar proof is used for when the E_i's are an decreasing sequence of events and if provided in the text by Ross. The fact that

$$
P\left(\lim_{n\to\infty} E_n\right) = \lim_{n\to\infty} P(E_n)
$$

says that P is a continuous set function.

A Paradox in Probability - One of Many in Probability

Suppose that we possess an infinitely large urn and an infinite collection of balls labeled ball number 1, number 2, number 3, and so on. Consider an experiment performed as follows: At 1 minute to 12 PM, balls numbered 1 through 10 are placed in the urn and ball number 10 is withdrawn, assuming the placing and withdrawing of the balls takes no time to occur. At 1/2 minute before 12 PM, balls numbered 11 through 20 are placed in the urn and ball number 20 is withdrawn. At 1/4 minute before 12 PM, balls numbered 21 through 30 are placed in the urn and ball number 30 is withdrawn. At 1/8 minute before 12 PM, balls numbered 31 through 40 are placed in the urn and ball number 40 is withdrawn, and this process is continued indefinitely. The problem of interest is to determine how many balls are in the urn at 12 PM. The answer to this is clearly that there are an infinite number of balls in the urn at 12 PM since the only balls in the urn at 12 PM are those labeled with numbers that are not evenly divisible by 10 and there are an infinite number of such numbers.

Now let us change the experiment slightly. Suppose that at 1 minute to 12 PM, balls numbered 1 through 10 are placed in the urn and ball number 1 is withdrawn, assuming the placing and withdrawing of the balls takes no time to occur. At 1/2 minute before 12 PM, balls numbered 11 through 20 are placed in the urn and ball number 2 is withdrawn. At 1/4 minute before 12 PM, balls numbered 21 through 30 are placed in the urn and ball number 3 is withdrawn. At 1/8 minute before 12 PM, balls numbered 31 through 40 are placed in the urn and ball number 4 is withdrawn, and this process is continued indefinitely. The problem of interest is to determine how many balls are in the urn at 12 PM. This time, the surprising answer is that the *urn is now empty* at 12 PM. This is because the ball that is labeled with the number n will be removed at $1/2^{n-1}$ minutes before 12 PM and hence every ball will be eventually removed from the urn by the time 12 PM comes along. Of course Zeno would argue that 12 PM never comes along in this case, but let us not worry about Zeno's paradox in reference to this problem.

Because for all n, the number of balls in the urn after the nth interchange is the same in both variations of the experiment (namely $9n$), most people are surprised that the two scenarios produce such different results in the limit as 12 PM comes along.

Of course, it is important to recognize that the reason the results are different is not because there is an actual paradox, or a mathematical contradiction, but rather because of the logic of the situation, and also that the surprise results because one's initial intuition (or common sense) when dealing with infinity is often flawed. After all, how often do you really ever deal with infinity?

This latter statement is not surprising for when the theory of infinity was developed by Georg Cantor is the second half of the 19th century, many other leading mathematicians of the day called it nonsensical and ridiculed Cantor for making claims such as there are some infinities bigger than other infinities and in fact there are an infinite number of different sizes to infinity. Wow!

We see from this example that the manner in which the balls are withdrawn seems to make a difference on how many balls are in the urn at 12 PM. For in the first case, only balls numbered 10, 20, 30, ..., are ever withdrawn whereas in the second case, all the balls are eventually withdrawn. Let us suppose that whenever a ball is to be withdrawn, that ball is *randomly selected* from among those present in the urn. That is, suppose at 1 minute to 12 PM, balls numbered 1 through 10 are placed in the urn and a ball is randomly selected from these 10 and withdrawn. At 1/2 minute to 12 PM, balls numbered 11 through 20 are placed in the urn and a ball is randomly selected from the 19 balls in the urn (remember that one was withdrawn earlier) and withdrawn. At at 1/4 minute to 12 PM, balls numbered 21 through 30 are placed in the urn and a ball is randomly selected from the 28 balls in the urn (remember that two were withdrawn earlier) and withdrawn, and so on. In this case, let us show (with probability *one*) that the urn will be empty at 12 PM.

To do this, let us first consider ball number 1. Define E_{1n} to be the event that the ball numbered 1 is still in the urn after the first n withdrawals have been made. Clearly

$$10 \times 19 \times 28 \times \cdots \times (9n+1)$$

equals the total number of ways that all the balls can be withdrawn and

$$(10-1) \times (19-1) \times (28-1) \times \cdots \times (9n+1-1) = 9 \times 18 \times 27 \times \cdots \times (9n)$$

equals the total number of ways in which the ball numbered 1 is not withdrawn. Assuming all are equally likely, we have

$$P(E_{1n}) = \frac{9 \times 18 \times 27 \times \cdots \times (9n)}{10 \times 19 \times 28 \times \cdots \times (9n+1)} = \prod_{k=1}^{n} \left(\frac{9k}{9k+1} \right)$$

which we may write as

$$P(E_{1n}) = \prod_{k=1}^{n} \left(\frac{1}{1+1/9k} \right) = \prod_{k=1}^{n} \left(1 + \frac{1}{9k} \right)^{-1}.$$

Now the event that ball number 1 is in the urn at 12 PM is just the event

$$E_{1\infty} = \bigcap_{n=1}^{\infty} E_{1n}$$

and since

$$P(E_{1\infty}) = P\left(\bigcap_{n=1}^{\infty} E_{1n}\right) = \lim_{n\to\infty} P(E_{1n}) = \lim_{n\to\infty} \prod_{k=1}^{n} \left(1 + \frac{1}{9k}\right)^{-1}$$

we have

$$P(E_{1\infty}) = \prod_{k=1}^{\infty} \left(1 + \frac{1}{9k}\right)^{-1} = \left\{\prod_{k=1}^{\infty} \left(1 + \frac{1}{9k}\right)\right\}^{-1}.$$

But

$$\begin{aligned}
\prod_{k=1}^{\infty} \left(1 + \frac{1}{9k}\right) &\geq \prod_{k=1}^{m} \left(1 + \frac{1}{9k}\right) \\
&= \left(1 + \frac{1}{9}\right)\left(1 + \frac{1}{18}\right)\left(1 + \frac{1}{27}\right)\cdots\left(1 + \frac{1}{9m}\right) \\
&\geq \frac{1}{9} + \frac{1}{18} + \frac{1}{27} + \cdots + \frac{1}{9m} \\
&= \frac{1}{9}\left(1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{m}\right) \\
&= \frac{1}{9}\sum_{k=1}^{m} \frac{1}{k}.
\end{aligned}$$

Thus we find that

$$\prod_{k=1}^{\infty} \left(1 + \frac{1}{9k}\right) \geq \frac{1}{9}\sum_{k=1}^{m} \frac{1}{k}$$

for all $m = 1, 2, 3, ...,$, and since

$$\lim_{m\to\infty} \sum_{k=1}^{m} \frac{1}{k} = +\infty$$

we see that

$$\prod_{k=1}^{\infty} \left(1 + \frac{1}{9k}\right) \geq +\infty$$

which says that

$$\prod_{k=1}^{\infty}\left(1+\frac{1}{9k}\right)=+\infty.$$

Then

$$P(E_{1\infty})=\{+\infty\}^{-1}=\frac{1}{+\infty}=0.$$

Therefore letting $E_{i\infty}$ being the event that ball number i is in the urn at 12 PM, we have shown that $P(E_{1\infty})=0$. Using a similar argument for ball numbered 2, we have $P(E_{2\infty})=0$ and so on for all the balls. This says that the probability that any ball is in the urn a 12 PM is zero and hence the probability that the urn is empty at 12 PM must be one. Very cool, right!

5. Probability as a Measure of Belief

Thus far we have interpreted the probability of an event of a given experiment as being a measure of how frequently the event will occur when the experiment is continually repeated. However, there are also other uses of the term probability. For instance, we have all heard such statements as "It is 90% probable that Shakespeare actually wrote *Hamlet*" or "The probability that Oswald acted alone assassinating Kennedy is 0.8". How are we to interpret these statements?

The most simple and natural interpretation is that the probabilities referred to measures of an individual's degree of belief in the statements that he or she is making. In other words, the individual making the forgoing statements is quite certain that Oswald acted alone and is even more certain that Shakespeare wrote *Hamlet*. This interpretation of probability as being a measure of the degree of one's belief is often referred to as the *personal* or *subjective* view of probability.

It seems logical to suppose that "a measure of the degree of one's belief" should satisfy all of the same axioms of probability. For example, if we are 70% certain that Shakespeare wrote *Julius Caesar* and 10% certain that it was actually Marlowe, then it is logical to suppose that we are 80% certain that it was either Shakespeare or Marlowe. Hence whether we interpret probability as a measure of belief or as a long-run frequency of occurrence, its mathematical properties remain unchanged.

Example #23: Playing the Horses

Suppose that in a 7-horse race, you believe that each of the first 2 horses has a 20% chance of winning, horses 3 and 4 each have a 15% chance, and the remaining 3 horses have a 10% chance each. Would it be better for you to wager at even money that the winner will be one the first three horses or to wager (again at even money) that the winner will be one of the horses 1, 5, 6 and 7?

To solve this, let 1 = 1 wins, 2 = 2 wins, ..., 7 = 7 wins. We are saying (based on our belief) that

$$P(1) = P(2) = 0.2 \quad , \quad P(3) = P(4) = 0.15 \quad \text{and} \quad P(5) = P(6) = P(7) = 0.1.$$

We want to compute

$$P(1 \cup 2 \cup 3) = P(1) + P(2) + P(3) = 0.2 + 0.2 + 0.15 = 0.55$$

and

$$P(1 \cup 5 \cup 6 \cup 7) = P(1) + P(5) + P(6) + P(7) = 0.2 + 0.1 + 0.1 + 0.1 = 0.5$$

showing that

$$P(1 \cup 2 \cup 3) > P(1 \cup 5 \cup 6 \cup 7)$$

and so you should wager that the winner will be one of the first three horses. ■

Removing Inconsistencies

Note that in supposing that a person's subjective probabilities are always consistent with the axioms of probability, we are dealing with an idealized rather than an actual person. For example, if we were to ask someone what he thought the chances were of: (a) rain today, (b) rain tomorrow, (c) rain both today and tomorrow, or (d) rain either today or tomorrow, it is quite possible that, after some deliberation, he might say 30%, 40%, 20% and 60% for (a), (b), (c) and (d), respectively. Unfortunately, such answers (or such subjective probabilities) are not consistent with the axioms of probabilities, since these answers imply

$$P(\text{rain today}) = 0.3 \quad , \quad P(\text{rain tomorrow}) = 0.4$$

and

$$P(\text{rain today } \cap \text{ rain tomorrow}) = 0.2$$

and

$$P(\text{rain today } \cup \text{ rain tomorrow}) = 0.6$$

but the axioms of probability says that we must have

$$\begin{aligned} P(\text{rain today } \cup \text{ rain tomorrow}) &= P(\text{rain today}) + P(\text{rain tomorrow}) \\ &\quad - P(\text{rain today } \cap \text{ rain tomorrow}) \end{aligned}$$

which would lead to

$$P(\text{rain today } \cup \text{ rain tomorrow}) = 0.3 + 0.4 - 0.2 = 0.5 \neq 0.6$$

as believed by the person. We would of course hope that after this was pointed out to the respondent, he would change his answers so that the equation

$$\begin{aligned} P(\text{rain today } \cup \text{ rain tomorrow}) &= P(\text{rain today}) + P(\text{rain tomorrow}) \\ &\quad - P(\text{rain today } \cap \text{ rain tomorrow}) \end{aligned}$$

holds true.