Jinguo Lian
Math437 Notes
January 25, 2016

1 Basic Principles of Sets

1.1.1 Sets

In Probability, we normally use sets to describe events.

Definition 1. *A set is a collection of objects, which are the elements of the set. If S is a set and x is an element of S, we write $x \in S$. If x is not an element of S,we write $x \notin S$. A set can have no elements, in which case it is called the empty set, denoted by $\emptyset$.*

Remark 2. *1. A is a* ***set*** *means that for any given object x, one can assert without dispute that either $x \in A$ (i.e., x belongs to A) or $x \notin A$ but not both.*

Example 3. *Which of the following is a well-defined set.*

(a) The collection of good movies. (answer: NO)

(b) The collection of right-handed individuals in Russellville. (Answer: Yes)

2. ***Set notation:***

   (a) *The first way is to list, without repetition, the elements of the set. For countable set like: $A = \{x_1, x_2, x_3, ...\}$ or the set of possible outcomes of a coin toss is $A = \{H, T\}$, where H stands for heads and T stands for tails.*

   (b) *The another way to represent a set is to describe a property that characterizes the elements of the set. Like $A = \{-3, 3\}$ or $A = \{x|x$ are solutions of $x^2 - 9 = 0\}$ or a uncountable set likes the set of all scalars x in the interval $[0, 1]$ can be written as $\{x|0 \leq x \leq 1\}$.*

3. ***Subsets, equal sets and universal set***

   (a) *If every element of a set A is also an element of a set B, we say that A is a subset of B , and we write $A \subset B$.*

   (b) *If $A \subset B$ and $B \subset A$, the two sets are equal, and we write $A = B$.*

   (c) *It is also expedient to introduce a universal set, denoted by Ω, which contains all objects that could conceivably be of interest in a particular context. Having specied the context in terms of a universal set Ω, we only consider sets S that are subsets of Ω.*

   (d) *The number of elements in a set is called the* ***cardinality of the set****. Use $n(A)$ to denote the cardinality of the set A. If A has a finite cardinality we say that A is a finite set. Otherwise, it is called infinite set. If A and B are two sets (finite or infinite) and there is a bijection from A to B(i.e., a one-to-one and onto function) then the two sets are said to have the same cardinality and we write $n(A) = n(B)$.*

1.1.2 Set Operations

Definition 4. *1. The complement of a set A, with respect to the universe Ω, is the set $\{x \in \Omega | x \notin A\}$ of all elements of Ω that do not belong to A, and is denoted by A^c. Note that $\Omega^c = \emptyset$.*

2. *The union of two sets A and B is the set of all elements that belong to A or B (or both), and is denoted by $A \cup B$ or $A \cup B = \{x | x \in A \text{ or } x \in B\}$.*

3. *The intersection of two sets A and B is the set of all elements that belong to both A and B , and is denoted by $A \cap B$ or $A \cap B = \{x | x \in A \text{ and } x \in B\}$.*

4. *we will have to consider the union or the intersection of several, even innitely many sets, dened in the obvious way. For example, if for every positive integer n, we are given a set A_n, then*

$$\cup_{n=1}^{\infty} A_n = A_1 \cup A_2 \cup \ = \{x | x \in A_n \text{ for some } n\}$$

$$\cap_{n=1}^{\infty} A_n = A_1 \cap A_2 \cap \ = \{x | x \in A_n \text{ for some } n\}$$

5. *If $A_i \bigcap A_j = \phi$, for all $i \neq j$ then we say that the sets in the collection $\{A_n\}_{n=1}^{\infty}$ are* ***pairwise disjoint****. A collection of sets is said to be a partition of a set S if the sets in the collection are pairwise disjoint and their union is S.*

1.1.3 The algebra of Sets

Set operations have several properties, which are elementary consequences of the denitions. Sets and the associated operations are easy to visualize in terms of Venn diagrams.

Remark 5. *1.* $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$

2. $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$

3. $(A \cap B)^c = (A^c \cup B^c)$

4. $(A \cup B)^c = (A^c \cap B^c)$

5. $(\cup_{n=1}^{\infty} A_n)^c = \cap_{n=1}^{\infty} A_n^c$

6. $(\cap_{n=1}^{\infty} A_n)^c = \cup_{n=1}^{\infty} A_n^c$

1.1.4 Practice problems

Example 6. *Express each of the following events in terms of the events A, B and C as well as the operations of complementation, union and intersection:*

1. *at least one of the events A, B, C occurs,*

2. *at most one of the events A, B, C occurs,*

3. *none of the events A, B, C occurs,*

4. *all three events A, B, C occur,*
5. *exactly one of the events A, B, C occurs,*
6. *events A and B occur, but not C,*
7. *either event A occurs or, if not, then B also does not occur.*

In each case draw the corresponding Venn diagram.

Solution:

1. $A \cup B \cup C$
2. $(A \cap B^c \cap C^c) \cup (A^c \cap B \cap C^c) \cup (A^c \cap B^c \cap C) \cup (A^c \cap B^c \cap C^c)$
3. $(A \cup B \cup C)^c = A^c \cap B^c \cap C^c$
4. $A \cap B \cap C$
5. $(A \cap B^c \cap C^c) \cup (A^c \cap B \cap C^c) \cup (A^c \cap B^c \cap C)$
6. $A \cap B \cap C^c$
7. $A \cup (A^c \cap B^c)$

Example 7. *Translate the following set-theoretic notation into event language. For example, $A \cup B$ means "A or B occurs".*

1. $A \cap B$
2. $A - B$
3. $A \cup B - A \cap B$
4. $A - (B \cup C)$
5. $A \subset B$
6. $A \cap B = \phi$

Solution:

1. *A and B occur*
2. *A occurs and B does not occur*
3. *A or B; but not both, occur*
4. *A occurs, and B and C do not occur*
5. *if A occurs, then B occurs but if B occurs then A need not occur.*
6. *if A occurs, then B does not occur or if B occurs then A does not occur*

Example 8. *A financial service company is studying a large pool of individuals who are potential clients. The company offers to sell its clients stocks, bonds and life insurance. The events of interest are the following:*

S: the individual owns stocks

B: the individual owns bonds

I: the individual has life insurance coverage

$I \cap (B \cup S) = (I \cap B) \cup (I \cap S)$: *insured and (owning bonds or stocks)*$\Leftrightarrow$ *(insured and owning bonds) or (insured and owing stocks).*

$I \cup (B \cap S) = (I \cup B) \cap (I \cup S)$: *insured or (owning bonds and stocks)*$\Leftrightarrow$ *(insured or owning bonds) and (insured or owing stocks).*

1.2 Basic Concepts of Probability

1.2.1 Basic definitions in the Sample Space

Definition 9.

1. *A random experiment is a process by which we observe something uncertain.*
2. *An outcome is a result of a random experiment.*
3. *Sample Space: The set of all possible outcomes. Denote by S.*
4. *Event: A subset of the sample space.*
5. *When we repeat a random experiment several times, we call each one of them a trial. Thus, a trial is a particular performance of a random experiment.*
6. *Our goal is to assign probability to certain events*

Example 10. *Random experiment: toss a coin; sample space:* $S = \{heads, tails\}$ *or as we usually write it,* $\{H, T\}$. *Random experiment: roll a die; sample space:* $S = \{1, 2, 3, 4, 5, 6\}$.

Remark 11.

1. *Union: If A and B are events, then* $A \cup B$ *occurs means A or B occur.*
2. *Intersection: If A and B are events,* $A \cap B$ *occurs means both A and B occur.*
3. *If* $A_1, A_2, ..., A_n$ *are events, then the event* $A_1 \cup A_2 \cup A_3 ... \cup A_n$ *occurs if at least one of* $A_1, A_2, ..., A_n$ *occurs.*
4. *The event* $A_1 \cap A_2 \cap A_3 ... \cap A_n$ *occurs if all of* $A_1, A_2, ..., A_n$ *occur.*
5. *It can be helpful to remember that the key words "or" and "at least" correspond to unions and the key words "and" and "all of" correspond to intersections.*

1.2.2 Axioms of Probability

Definition 12. *Axioms of Probability:*

1. *Axiom 1: For any event A,* $P(A) \geq 0$.
2. *Axiom 2: Probability of the sample space S is* $P(S) = 1$.
3. *Axiom 3: If* $A_1, A_2, A_3, ...$ *are disjoint events, then* $P(A_1 \cup A_2 \cup A_3 ...) = P(A_1) + P(A_2) + P(A_3) + ...$

Remark 13. 1. *For any event A,* $P(A^c) = 1P(A)$.

2. *The probability of the empty set is zero, i.e.,* $P(\emptyset) = 0$.
3. *For any event A,* $P(A) \leq 1$.
4. $P(A - B) = P(A) - P(A \cap B)$.
5. $P(A \cup B) = P(A) + P(B)P(A \cap B)$, *(inclusion-exclusion principle for n=2).*
6. *If* $A \subset B$ *then* $P(A) \leq P(B)$.
7. $P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$

Example 14. *In a presidential election, there are four candidates. Call them A, B, C, and D. Based on our polling analysis, we estimate that A has a 20 percent chance of winning the election, while B has a 40 percent chance of winning. What is the probability that A or B win the election?*

Solution: *Notice that the events that A wins, B wins, C wins, and D wins are disjoint*

$$P(AwinsorBwins) = P(\{Awins\} \cup \{Bwins\}) = P(\{Awins\}) + P(\{Bwins\}) = 0.6$$

1.2.3 Discrete Probability Models-infinite countable or finite sample space with unequally likely outcomes

Consider a sample space S. If S is a countable set, this refers to a discrete probability model. In this case, since S is countable, we can list all the elements in S: $S = \{s_1, s_2, s_3, \cdots\}$.

If $A \subset S$ is an event, then A is also countable, and by the third axiom of probability we can write

$$P(A) = P(\cup_{s_j \in A}\{s_j\}) = \sum_{s_j \in A} P(s_j)$$

Thus in a countable sample space, to find probability of an event, all we need to do is sum the probability of individual elements in that set.

Example 15. *I play a gambling game in which I will win k-2 dollars with probability* $\frac{1}{2^k}$ *for any* $k \in N$, *that is,*

1. *with probability 1/2, I lose 1 dollar;*

2. *with probability 1/4, I win 0 dollar;*
3. *with probability 1/8, I win 1 dollar;*
4. *with probability 1/16, I win 2 dollars;*
5. *with probability 1/32, I win 3 dollars;*
6. *...*

What is the probability that I win more than or equal to 1 dollar and less than 4 dollars? What is the probability that I win more than 2 dollars?

hint: *Note: Here we have used the geometric series sum formula. In particular, for any a,x $\in R$, if $|x| < 1$ we have*

$$a + ax + ax^2 + ax^3 + \cdots + ax^{n-1} = \sum_{k=0}^{n-1} ax^k = \frac{a(1-x^n)}{1-x} \to \frac{a}{1-x}$$

1.2.4 Discrete Probability Models-Finite Sample Spaces with Equally Likely Outcomes

An important special case of discrete probability models is when we have a finite sample space S, where each outcome is equally likely, i.e., $S = \{s_1, s_2, ..., s_N\}$, where $P(s_i) = P(s_j)$ for all $i, j \in 1, 2, ..., N$. Rolling a fair die is an instance of such a probability model. Since all outcomes are equally likely, we must have $P(s_i) = \frac{1}{N}$, for all $i \in \{1, 2, \cdots, N\}$. For all $P(s_i) = \frac{1}{N}$, for all $i \in \{1, 2, \cdots, N\}$. In such a model, if A is any event with cardinality n(A)=M, we can write

$$P(A) = \sum_{s_j \in A} P(s_j) = \sum_{s_j \in A} \frac{1}{N} = \frac{M}{N} = \frac{n(A)}{n(S)}$$

. Thus, finding probability of A reduces to a counting problem in which we need to count how many elements are in A and S.

Example 16. *I roll a fair die twice and obtain two numbers: X_1= result of the first roll, and X_2= result of the second roll. Write down the sample space S, and assuming that all outcomes are equally likely (because the die is fair), find the probability of the event A defined as the event that $X_1 + X_2 = 8$.*

Solution: *$n(S) = 6 \times 6 = 36$, $A = \{(2, 6), (3, 5), (4, 4), (5, 3), (6, 2)\}$, so $n(A) = 5$, $P(A) = \frac{n(A)}{n(S)} = \frac{5}{36}$.*

Remark 17. *In a finite sample space S, where all outcomes are equally likely, the probability of any event A can be found by $P(A) = \frac{n(A)}{n(S)}$.*

Definition 18. *The **Cartesian product** of two sets A and B is the set $A \times B = \{(a, b) | a \in A, b \in B\}$. The idea can be extended to products of any number of sets.*

Remark 19. *If S is the sample space then $S = S_1 \times S_2 \times ... \times S_n$, where $S_i, 1 \leq i \leq n$, is the set consisting of the two outcomes H=head and T = tail. The following theorem is a tool for finding the cardinality of the Cartesian product of two finite sets. Given two finite sets A and B. Then $n(A \times B) = n(A) \times n(B)$.*

Example 20. *What is the total number of outcomes of tossing a fair coin n times.*

Solution: *If S is the sample space then $S = S_1 \times S_2 \times ... \times S_n$ where $S_i, 1 \leq i \leq n$, is the set consisting of the two outcomes H=head and T = tail. so, $n(S) = 2^n$.*

1.3 Combinatorics

1.3.1 Counting

Remember that for a finite sample space S with equally likely outcomes, the probability of an event A is given by

$$P(A) = \frac{n(A)}{n(S)} = \frac{M}{N}$$

Thus, finding probability of A reduces to a counting problem in which we need to count how many elements are in A and S. In this section, we will discuss ways to count the number of elements in a set in an efficient manner. Counting is an area of its own and there are books on this subject alone. Here we provide a basic introduction to the material that is usually needed in probability. Almost everything that we need about counting is the result of the multiplication principle. We previously saw the multiplication principle when we were talking about Cartesian products.

Definition 21. *Multiplication Principle Suppose that we perform r experiments such that the k^{th} experiment has n_k possible outcomes, for $k = 1, 2, ..., r$. Then there are a total of $n_1 \times n_2 \times n_3 \times ... \times n_r$ possible outcomes for the sequence of r experiments.*

Here, we would like to provide some general terminology for the counting problems that show up in probability to make sure that the language that we use is precise and clear.

Definition 22.

1. ***Sampling:*** *sampling from a set means choosing an element from that set. We often draw a sample at random from a given set in which each element of the set has equal chance of being chosen.*

2. ***With or without replacement:*** *usually we draw multiple samples from a set. If we put each object back after each draw, we call this sampling with replacement. In this case a single object can be possibly chosen multiple times. For example, if $A = \{a_1, a_2, a_3, a_4\}$ and we pick 3 elements with replacement, a possible choice might be (a_3, a_1, a_3). Thus "with replacement" means "repetition is allowed." On the other hand, if repetition is not allowed, we call it sampling without replacement.*

3. ***Ordered or unordered:*** *If ordering matters (i.e.: $a_1, a_2, a_3 \neq a_2, a_3, a_1$), this is called ordered sampling. Otherwise, it is called unordered.*

Thus when we talk about sampling from sets, we can talk about four possibilities.

Remark 23. *Let's summarize the formulas for the four categories of sampling. Assuming that we have a set with n elements, and we want to draw k samples from the set, then the total number of ways we can do this is given by the following table.*

ordered sampling with replacement: n^k

ordered sampling without replacement: $P_k^n = \frac{n!}{(n-k)!}$

unordered sampling without replacement: $\binom{n}{k} = \frac{n!}{k!(n-k)!}$

unordered sampling with replacement: $\binom{n+k-1}{k}$

We will discuss each of these in detail and indeed will provide a formula for each. Nevertheless, the best approach here is to understand how to derive these formulas. You do not actually need to memorize them if you understand the way they are obtained.

1.3.2 Ordered Sampling with Replacement

Example 24. *Here we have a set with n elements (e.g.: $A = \{1, 2, 3, ...n\}$), and we want to draw k samples from the set such that ordering matters and repetition is allowed. For example, if $A = \{1, 2, 3\}$ and k=2, there are 9 different possibilities:*

$$(1, 1), (1, 2), (1, 3), (2, 1), (2, 2), (2, 3), (3, 1), (3, 2), (3, 3)$$

In general, we can argue that there are k positions in the chosen list: (Position 1, Position 2, ..., Position k). There are n options for each position. Thus, when ordering matters and repetition is allowed, the total number of ways to choose k objects from a set with n elements is

$$n \times n \times ... \times n = n^k$$

Note that this is a special case of the multiplication principle where there are k "experiments" and each experiment has n possible outcomes.

1.3.3 Ordered Sampling without Replacement:Permutations

Example 25. *Consider the same setting as above, but now repetition is not allowed. For example, if $A = \{1, 2, 3\}$ and k=2, there are 6 different possibilities:*

$$(1, 2), (1, 3), (2, 1), (2, 3), (3, 1), (3, 2)$$

. In general, we can argue that there are k positions in the chosen list: (Position 1, Position 2, ..., Position k). There are n options for the first position, (n-1) options for the second position (since one element has already been allocated to the first position and cannot be chosen here), (n-2) options for the third position, ... (n-k+1) options for the kth position. Thus, when ordering matters and repetition is not allowed, the total number of ways to choose k objects from a set with n elements is

$$n \times (n-1) \times ... \times (n-k+1)$$

Any of the chosen lists in the above setting (choose k elements, ordered and no repetition) is called a k-permutation of the elements in set A. We use the following notation to show the number of k-permutations of an n-element set:

$$P_n^k = n \times (n-1) \times ... \times (n-k+1)$$

Note that if k is larger than n, then $P_n^k = 0$. This makes sense, since if $k > n$ there is no way to choose k distinct elements from an n-element set. Let's look at a very famous problem, called the birthday problem.

Example 26. *If k people are at a party, what is the probability that at least two of them have the same birthday? Suppose that there are n=365 days in a year and all days are equally likely to be the birthday of a specific person.*

Solution: *Let A be the event that at least two people have the same birthday. First note that if $k > n$, then $P(A) = 1$, so, let's focus on the more interesting case where $k \leq n$. Again, the phrase "at least" suggests that it might be easier to find the probability of the complement event, $P(A^c)$. This is the event that no two people have the same birthday, and we have*

$$P(A) = 1 - \frac{n(A^c)}{n(S)}$$

Thus, to solve the problem it suffices to find $n(A^c)$ and n(S). Let's first find n(S). What is the total number of possible sequences of birthdays of k people? Well, there are n=365 choices for the first person, n=365 choices for the second person,... n=365 choices for the k^th person. Thus there are n^k possibilities. This is, in fact, an ordered sampling with replacement problem, and as we have discussed, the answer should be n^k (here we draw k samples, birthdays, from the set 1,2,...,n=365). Now let's find $n(A^c)$. If no birthdays are the same, this is similar to finding n(S) with the difference that repetition is not allowed, so we have

$$n(A^c) = P_n^k = n \times (n-1) \times ... \times (n-k+1)$$

You can see this directly by noting that there are n=365 choices for the first person, n1=364 choices for the second person,..., n-k+1 choices for the k^th person. Thus the probability of A can be found as

$$P(A) = 1 - \frac{n(A^c)}{n(S)} = 1 - \frac{P_n^k}{n^k}$$

Remark 27. *Permutations of n elements: An n-permutation of n elements is just called a permutation of those elements. In this case, k=n and we have*

$$P_n^n = n \times (n-1) \times ... \times (n-n+1) = n \times (n-1) \times ... \times 1$$

,

which is denoted by n!, pronounced "n factorial". Thus n! is simply the total number of permutations of n elements, i.e., the total number of ways you can order n different objects. To make our formulas consistent, we define 0!=1.

The number of k-permutations of n distinguishable objects is given by

$$P_n^k = \frac{n!}{(n-k)!}, \textit{ for } 0 \leq k \leq n$$

1.3.4 Unordered Sampling without Replacement:Combinations

Here we have a set with n elements, e.g., $A = \{1, 2, 3, ...n\}$ and we want to draw k samples from the set such that ordering does not matter and repetition is not allowed. Thus, we basically want to choose a k-element subset of A, which we also call a k-combination of the set A. For example if $A = \{1, 2, 3\}$ and k=2, there are 3 different possibilities:

$$\{1, 2\}, \{1, 3\}, \{2, 3\}$$

We show the number of k-element subsets of A by $\binom{n}{k}$. This is read "n choose k." A typical scenario here is that we have a group of n people, and we would like to choose k of them to serve on a committee. A simple way to find $\binom{n}{k}$ is to compare it with P_n^k. Note that the difference between the two is ordering. In fact, for any k-element subset of $A = \{1, 2, 3, ...n\}$, we can order the elements in k! ways, thus we can write

$$P_n^k = \binom{n}{k} \times k!$$

Therefore,

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

Note that if k is an integer larger than n, then $\binom{n}{k}$=0. This makes sense, since if k¿n there is no way to choose k distinct elements from an n-element set.

Definition 28. *The number of k-combinations of an n-element set is given by*

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}, \text{ for } 0 \le k \le n$$

.

Remark 29. $\binom{n}{k}$ *is also called the binomial coefficient. This is because the coefficients in the binomial theorem are given by* $\binom{n}{k}$. *In particular, the binomial theorem states that for an integer* $n \ge 0$*, we have*

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}$$

Example 30. *I choose 3 cards from the standard deck of cards. What is the probability that these cards contain at least one ace?*

Solution: *Again the phrase "at least" suggests that it might be easier to first find P(Ac), the probability that there is no ace. Here the sample space contains all possible ways to choose 3 cards from 52 cards, thus*

$$n(S) = \binom{52}{3}$$

There are 524=48 non-ace cards, so we have

$$n(A^c) = \binom{48}{3}$$

. *Thus*

$$P(A) = 1 - \frac{\binom{48}{3}}{\binom{52}{3}}$$

Example 31. *How many distinct sequences can we make using 3 letter "A"s and 5 letter "B"s? (AAABBBBB, AABABBBB, etc.)*

Solution:

You can think of this problem in the following way. You have 3+5=8 positions to fill with letters A or B. From these 8 positions, you need to choose 3 of them for As. Whatever is left will be filled with Bs. Thus the total number of ways is $\binom{8}{3}$. *Now, you could have equivalently chosen the locations for Bs, so the answer would have been* $\binom{8}{5}$. *Thus, we conclude that* $\binom{8}{3} = \binom{8}{3}$. *More generally, we have* $\binom{n}{k} = \binom{n}{n-k}$

Remark 32. *We have*

$$\sum_{k=0}^{n} \binom{n}{k} = 2^n$$

For $0 \leq k < n$, *we have*

$$\binom{n+1}{k+1} = \binom{n}{k+1} + \binom{n}{k}$$

Now, we are ready to discuss an important class of random experiments that appear frequently in practice. First, we define Bernoulli trials and then discuss the binomial distribution. A Bernoulli Trial is a random experiment that has two possible outcomes which we can label as "success" and "failure," such as

Definition 33. *You toss a coin. The possible outcomes are "heads" and "tails." You can define "heads" as success and "tails" as "failure" here.*

You take a pass-fail test. The possible outcomes are "pass" and "fail."

We usually denote the probability of success by p and probability of failure by q=1p. If we have an experiment in which we perform n independent Bernoulli trials and count the total number of successes, we call it a binomial experiment. For example, you may toss a coin n times repeatedly and be interested in the total number of heads.

Binomial Formula: For n independent Bernoulli trials where each trial has success probability p, the probability of k successes is given by

$$P(k) = \binom{n}{k} p^k (1-p)^{n-k}$$

Example 34. *Ten people have a potluck. Five people will be selected to bring a main dish, three people will bring drinks, and two people will bring dessert. How many ways they can be divided into these three groups?*

Solution: *We can solve this problem in the following way. First, we can choose 5 people for the main dish. This can be done in* $\binom{10}{5}$ *ways. From the remaining 5 people, we then choose 3 people for drinks, and finally the remaining 2 people will bring desert. Thus, by the multiplication principle, the total number of ways is given by*

$$\binom{10}{5}\binom{5}{3}\binom{2}{2} = \frac{10!}{5!5!} \times \frac{5!}{3!2!} \times \frac{2!}{2!0!} = \frac{10!}{5!3!2!}$$

Remark 35. *This argument can be generalized for the case when we have n people and would like to divide them to r groups. The number of ways in this case is given by the multinomial coefficients. In particular, if $n = n_1 + n_2 + ... + n_r$, where all ni0 are integers, then the number of ways to divide n distinct objects to r distinct groups of sizes $n_1, n_2, ..., n_r$ is given by*

$$\binom{n}{n_1, n_2, ..., n_r} = \frac{n!}{n_1!n_2!...n_r!}$$

We can also state the general format of the binomial theorem, which is called the multinomial theorem:

$$(x_1 + x_2 + ... + x_r)^n = \sum_{n_1+n_2+...+n_r=n} \binom{n}{n_1, n_2, , n_r} x_1^{n_1} x_2^{n_2} ... x_r^{n_r}$$

Remark 36. *We now state the general form of the multinomial formula. Suppose that an experiment has r possible outcomes, so the sample space is given by*

$$S = \{s_1, s_2, ..., s_r\}$$

Also suppose that $P(s_i) = p_i$ for $i = 1, 2, ..., r$. Then for $n = n_1 + n_2 + ... + n_r$ independent trials of this experiment, the probability that each s_i appears ni times is given by

$$\binom{n}{n_1, n_2, ..., n_r} p_1^{n_1} p_2^{n_2} ... p_r^{n_r} = \frac{n!}{n_1!n_2!...n_r!} p_1^{n_1} p_2^{n_2} ... p_r^{n_r}$$

1.3.5 Unordered Sampling with Replacement

Among the four possibilities we listed for ordered/unordered sampling with/without replacement, unordered sampling with replacement is the most challenging one. Suppose that we want to sample from the set $A = \{a_1, a_2, ..., a_n\}$ k times such that repetition is allowed and ordering does not matter. For example, if $A = \{1, 2, 3\}$ and k=2, then there are 6 different ways of doing this

$$(1, 1), (1, 2), (1, 3), (2, 2), (2, 3), (3, 3)$$

How can we get the number 6 without actually listing all the possibilities? One way to think about this is to note that any of the pairs in the above list can be represented by the number of 1's, 2's and 3's it contains. That is, if x_1 is the number of ones, x_2 is the number of twos, and x_3 is the number of threes, we can equivalently represent each pair by a vector (x_1, x_2, x_3), i.e.,

$$\{1, \}) \to (x_1, x_2, x_3) = (2, 0, 0)$$

$$\{1, 2\} \to (x_1, x_2, x_3) = (1, 1, 0)$$

$$\{1, 3\} \to (x_1, x_2, x_3) = (1, 0, 1)$$

$$\{2, 2\} \to (x_1, x_2, x_3) = (0, 2, 0)$$

$$\{2, 3\} \to (x_1, x_2, x_3) = (0, 1, 1)$$

$$\{3, 3\} \to (x_1, x_2, x_3) = (0, 0, 2)$$

Note that here $x_i \geq 0$ are integers and $x_1 + x_2 + x_3 = 2$. Thus, we can claim that the number of ways we can sample two elements from the set $A = \{1, 2, 3\}$ such that ordering does not matter and repetition is allowed is the same as solutions to the following equation $x_1 + x_2 + x_3 = 2$, where $x_i \in \{0, 1, 2\}$. This is an interesting observation and in fact using the same argument we can make the following statement for general k and n.

Theorem 37. *The number of distinct solutions to the equation*

$$x_1 + x_2 + ... + x_n = k, \ \textit{where } x_i \in \{0, 1, 2, 3, ...k\}$$

is equal to

$$\binom{n+k-1}{k} = \binom{n+k-1}{n-1}$$

Example 38. *Ten passengers get on an airport shuttle at the airport. The shuttle has a route that includes 5 hotels, and each passenger gets off the shuttle at his/her hotel. The driver records how many passengers leave the shuttle at each hotel. How many different possibilities exist?*

Solution:

Let x_i be the number of passengers that get off the shuttle at Hotel i. Then we have

$$x_1 + x_2 + x_3 + x_4 + x_5 = 10, \ \textit{where } x_i \in \{0, 1, 2, 3, ...10\}$$

Thus, the number of solutions is

$$\binom{5+10-1}{10} = \binom{5+10-1}{51} = \binom{14}{4}$$

1.4 Basic Concepts of Probability

1.4.1 Conditional Probability

Definition 39. *If A and B are two events in a sample space S, then the conditional probability of A given B is defined as*

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \ \textit{when } P(B) > 0$$

.

Remark 40. *It is important to note that conditional probability itself is a probability measure, so it satisfies probability axioms. In particular,*

1. *Axiom 1: For any event A, $P(A|B) \geq 0$.*
2. *Axiom 2: Conditional probability of B given B is 1, i.e., $P(B|B) = 1$.*
3. *Axiom 3: If $A_1, A_2, A_3, ...$ are disjoint events, then $P(A_1 \cup A_2 \cup A_3 ...|B) = P(A_1|B) + P(A_2|B) + P(A_3|B) + ...$*

Remark 41. *In fact, all rules that we have learned so far can be extended to conditional probability. For three events, A, B, and C, with* $P(C) > 0$*, we have*

1. $P(A^c|C) = 1 - P(A|C)$;
2. $P(\emptyset|C) = 0$;
3. $P(A|C) \leq 1$;
4. $P(A - B|C) = P(A|C) - P(A \cap B|C)$;
5. $P(A \cup B|C) = P(A|C) + P(B|C) - P(A \cap B|C)$;
6. *if* $A \subset B$ *then* $P(A|C) \leq P(B|C)$.

Example 42. *I roll a fair die twice and obtain two numbers* X_1= *result of the first roll and* X_2= *result of the second roll. Given that I know* $X_1 + X_2$=7*, what is the probability that* $X_1 = 4$ *or* $X_2 = 4$*?*

Solution: *Let A be the event that* $X_1 = 4$ *or* $X_2 = 4$ *and B be the event that* $X_1 + X_2 = 7$*. We are interested in* $P(A|B)$*, so we can use* $P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{\frac{2}{36}}{\frac{6}{36}} = \frac{1}{3}$.

Remark 43. *Chain rule for conditional probability:* $P(A \cap B) = P(A)P(B|A) = P(B)P(A|B)$

More general,Chain rule for conditional probability:

$$P(A_1 \cap A_2 \cap \cdots \cap A_n) = P(A_1)P(A_2|A_1)P(A_3|A_2, A_1) \cdots P(A_n|A_{n-1}A_{n-2} \cdots A_1)$$

Example 44. *In a factory there are 100 units of a certain product, 5 of which are defective. We pick three units from the 100 units at random. What is the probability that none of them are defective?*

Solution: *Let us define* A_i *as the event that the ith chosen unit is not defective, for i=1,2,3. We are interested in* $P(A_1 \cap A_2 \cap A_3)$*. Note that* $P(A_1) = \frac{95}{100}$, $P(A_2|A_1) = \frac{94}{99}$ *and* $P(A_3|A_2, A_1) = \frac{93}{98}$*. Then* $P(A_1 \cap A_2 \cap A_3) = \frac{95}{100}\frac{94}{99}\frac{93}{98} = 0.8560$.

Definition 45. *Two events A and B are independent if* $P(A \cap B) = P(A)P(B) \Rightarrow P(A|B) = P(A)$*. This means that knowing that B has occurred does not change our belief about the probability of A.*

Warning! One common mistake is to confuse independence and being disjoint. These are completely different concepts. When two events A and B are disjoint it means that if one of them occurs, the other one cannot occur, i.e., $A \cap B = \emptyset$. Thus, event A usually gives a lot of information about event B which means that they cannot be independent. See this theorem.

Theorem 46. *Consider two events A and B, with* $P(A) \neq 0$ *and* $P(B) \neq 0$*. If A and B are disjoint, then they are not independent.*

See the following example,

Example 47. *I pick a random number from $\{1, 2, 3, ..., 10\}$, and call it N. Suppose that all outcomes are equally likely. Let A be the event that N is less than 7, and let B be the event that N is an even number. Are A and B independent?*

Solution: *We have $A = \{1, 2, 3, 4, 5, 6\}$, $B = \{2, 4, 6, 8, 10\}$, and $A \cap B = \{2, 4, 6\}$. Then $P(A) = 0.6, P(B) = 0.5, P(A \cap B) = 0.3$. Therefore, $P(A \cap B) = P(A)P(B)$, so A and B are independent.*

Remark 48. *If $P(A_i \cap A_j) = P(A_i)P(A_j)$, for all $i, j \in \{1, 2, \cdots, n\}$, then $P(A_i \cap A_j \cap A_k) = P(A_i)P(A_j)P(A_k)$, for all $i, j, k \in \{1, 2, \cdots, n\}$*

Example 49. *I toss a coin repeatedly until I observe the first tails at which point I stop. Let X be the total number of coin tosses. Find $P(X = 5)$? Since all the coin tosses are independent, we can write $P(HHHHT) = P(H)P(H)P(H)P(H)P(T) =$*
$frac132$.

Theorem 50. *If A and B are independent then*

1. *A and B^c are independent,*
2. *A^c and B are independent,*
3. *A^c and B^c are independent.*

Theorem 51. *If $A_1, A_2, ..., A_n$ are independent then $P(A_1 \cup A_2 \cup ... \cup A_n) = 1(1P(A_1))(1P(A_2))...(1P(A_n))$.*

Remark 52. *If $A \cap B = \emptyset$ is disjoint, then $P(A \cup B) = P(A) + P(B)$.*

If A and B are independent, then $P(A|B) = P(A)$, $P(B|A) = P(B)$ and $P(A \cap B) = P(A)P(B)$.

1.4.2 Law of Total Probability

Definition 53. *Law of Total Probability:*

If $B_1, B_2, B_3, ...$ is a partition of the sample space S, then for any event A we have

$$P(A) = \sum_i P(A \cap B_i) = \sum_i P(A|B_i)P(B_i)$$

Example 54. *I have three bags that each contain 100 marbles:*

Bag 1 has 75 red and 25 blue marbles;
Bag 2 has 60 red and 40 blue marbles;
Bag 3 has 45 red and 55 blue marbles.

I choose one of the bags at random and then pick a marble from the chosen bag, also at random. What is the probability that the chosen marble is red?

Solution: *Let R be the event that the chosen marble is red. Let Bi be the event that I choose Bag i. We already know that*

$$P(R|B_1) = 0.75$$

,

$$P(R|B_2) = 0.60$$

,

$$P(R|B_3) = 0.45$$

We choose our partition as B_1, B_2, B_3. Note that this is a valid partition because, firstly, the B_i's are disjoint (only one of them can happen), and secondly, because their union is the entire sample space as one the bags will be chosen for sure, i.e., $P(B_1 \cup B_2 \cup B_3) = 1$. Using the law of total probability, we can write

$$\begin{aligned} P(R) &= P(R|B_1)P(B_1) + P(R|B_2)P(B_2) + P(R|B_3)P(B_3) \\ &= (0.75)\frac{1}{3} + (0.60)\frac{1}{3} + (0.45)\frac{1}{3} = 0.60 \end{aligned}$$

1.4.3 Bayes' Rule

Now we are ready to state one of the most useful results in conditional probability: Bayes' rule. Suppose that we know P(A—B), but we are interested in the probability P(B—A). Using the definition of conditional probability, we have

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

Dividing by P(A), we obtain

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}$$

, which is the famous Bayes' rule.

Definition 55. *Bayes' Rule*

For any two events A and B, where $P(A) \neq 0$, we have

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}$$

If $B_1, B_2, B_3, \ldots$ form a partition of the sample space S, and A is any event with $P(A) \neq 0$, we have

$$P(B_j|A) = \frac{P(A|B_j)P(B_j)}{\sum_i P(A|B_i)P(B_i)}$$

Example 56. *A certain disease affects about 1 out of 10,000 people. There is a test to check whether the person has the disease. The test is quite accurate. In particular, we know that the probability that the test result is positive (suggesting the person has the disease), given that the person does not have the disease, is only 2 percent; the probability that the test result is negative (suggesting the person does not have the disease), given that the person has the disease, is only 1 percent. A random person gets tested for the disease and the result comes back positive. What is the probability that the person has the disease?*

Solution: *Let D be the event that the person has the disease, and let T be the event that the test result is positive. We know*

$$P(D) = \frac{1}{10,000}$$

,

$$P(T|D^c) = 0.02$$

,

$$P(T^c|D) = 0.01$$

What we want to compute is $P(D|T)$. Again, we use Bayes' rule:

$$P(D|T) = \frac{P(T|D)P(D)}{P(T|D)P(D) + P(T|D^c)P(D^c)}$$

$$= \frac{(1-0.01)\times 0.0001}{(1-0.01)\times 0.0001 + 0.02\times(1-0.0001)}$$

$$= 0.0049$$

This means that there is less than half a percent chance that the person has the disease.

1.4.4 Conditional Independence

As we mentioned earlier, almost any concept that is defined for probability can also be extended to conditional probability. Remember that two events A and B are independent if $P(A\cap B) = P(A)P(B)$,or equivalently, $P(A|B) = P(A)$. We can extend this concept to conditionally independent events. In particular,

Definition 57. *Two events A and B are conditionally independent given an event C with $P(C) > 0$ if*

$$P(A\cap B|C) = P(A|C)P(B|C)$$

or

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

Example 58. *A box contains two coins: a regular coin and one fake two-headed coin ($P(H) = 1$). I choose a coin at random and toss it twice. Define the following events.*

A= First coin toss results in an H.

B= Second coin toss results in an H.

C= Coin 1 (regular) has been selected.

Find $P(A|C), P(B|C), P(A\cap B|C), P(A), P(B)$, and $P(A\cap B)$. Note that A and B are NOT independent, but they are conditionally independent given C.

Solution: *We have $P(A|C) = P(B|C) = \frac{1}{2}$. Also, given that Coin 1 is selected, we have $P(A\cap B|C) = \frac{1}{2}\times\frac{1}{2} = \frac{1}{4}$. To find $P(A), P(B)$, and $P(A\cap B)$, we use the law of total probability:*

$$P(A) = P(A|C)P(C) + P(A|C^c)P(C^c) = \frac{1}{2}\times\frac{1}{2} + 1\times\frac{1}{2} = \frac{3}{4}$$

Similarly, $P(B) = \frac{3}{4}$. For $P(A\cap B)$, we have

$$P(A\cap B) = P(A\cap B|C)P(C) + P(A\cap B|C^c)P(C^c)$$
$$= P(A|C)P(B|C)P(C) + P(A|C^c)P(B|C^c)P(C^c)$$

(by conditional independence of A and B)

$$= \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} + 1 \times 1 \times \frac{1}{2} = \frac{5}{8}$$

As we see, $P(A \cap B) = 58 \neq P(A)P(B) = \frac{9}{16}$, which means that A and B are not independent.

1.4.5 Practice problems

Example 59. *You purchase a certain product. The manual states that the lifetime T of the product, defined as the amount of time (in years) the product works properly until it breaks down, satisfies $P(T \geq t) = e^{t/5}$, for all $t \geq 0$. For example, the probability that the product lasts more than (or equal to) 2 years is $P(T \geq 2) = e^{2/5} = 0.6703$. I purchase the product and use it for two years without any problems. What is the probability that it breaks down in the third year?*

Solution: *Let A be the event that a purchased product breaks down in the third year. Also, let B be the event that a purchased product does not break down in the first two years. We are interested in $P(A|B)$. We have*

$$P(B) = P(T \geq 2) = e^{2/5}$$

We also have $P(A) = P(2 \geq T \leq 3) = P(T \geq 2)P(T \geq 3) = e^{2/5}e^{3/5}$.

Finally, since $A \subset B$, we have $A \cap B = A$. Therefore,

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)}{P(B)} = \frac{e^{2/5} - e^{3/5}}{e^{2/5}} = 0.1813$$

Example 60. *You toss a fair coin three times:*

What is the probability of three heads, HHH?

What is the probability that you observe exactly one heads?

Given that you have observed at least one heads, what is the probability that you observe at least two heads?

Solution: *We assume that the coin tosses are independent.*

$$P(HHH) = P(H)P(H)P(H) = 0.5^3 = 1/8$$

. To find the probability of exactly one heads, we can write

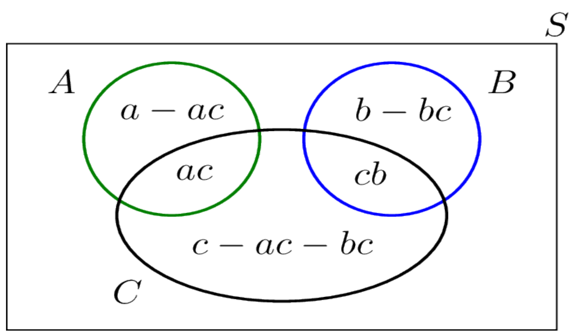
$$\begin{aligned} P(Oneheads) &= P(HTT \cup THT \cup TTH) \\ &= P(HTT) + P(THT) + P(TTH) \\ &= 1/8 + 1/8 + 1/8 \\ &= 3/8 \end{aligned}$$

Given that you have observed at least one heads, what is the probability that you observe at least two heads? Let A_1 be the event that you observe at least one heads, and A_2 be the *event that you observe at least two heads. Then $A_1 = STTT$, and $P(A_1) = 7/8$, $A_2 = HHT, HTH, THH, HHH$, and $P(A_2) = 4/8$. Thus, we can write*

$$P(A_2|A_1) = \frac{P(A_2 \cap A_1)}{P(A_1)} = \frac{P(A_2)}{P(A_1)} = (4/8)(8/7) = 4/7$$

Example 61. *For three events A, B, and C, we know that*
A and C are independent,
B and C are independent,
A and B are disjoint,
$P(A \cup C) = 2/3$,$P(B \cup C) = 3/4$,$P(A \cup B \cup C) = 11/12$ Find $P(A), P(B)$, and $P(C)$.

Solution: *We can use the Venn diagram in Figure to better visualize the events in this problem. We assume P(A)=a,P(B)=b, and P(C)=c. Note that the assumptions about independence and disjointness of sets are already included in the figure*



$P(A) = a, P(B) = b, P(C) = c$

Now we can write

$$P(A \cup C) = a + c - ac = 2/3$$

;

$$P(B \cup C) = b + c - bc = 3/4$$

;

$$P(A \cup B \cup C) = a + b + c - ac - bc = 11/12$$

. *By subtracting the third equation from the sum of the first and second equations, we immediately obtain $c = 1/2$, which then gives $a = 1/3$ and $b = 1/2$.*