Southwestern University Of Finance And Economics spring, 2018
School of Finance Lecture 1
Scribe: *Haoming Wang* Lecturer: *Prof. Yongmiao Hong*

Probability: Foundation of Probability Theory

Key word: Random Experiment, Concepts of Probability, Set Theory, Fundamental Probability Laws, Methods of Counting, Conditional Probability, Multiplication Rules, Bayes's Theorem, Independence

1 Random Experiment

Definition 1.1: Random Experiment

A **random experiment** is a mechanism which has at least two possible outcomes. When a random experiment is performed, one and only one outcome will occur, but which outcome to occur is unknown in advance.

There are two essential elements of a random experiment:

- the set of all possible outcomes;
- the likelihood with which each outcome will occur.

From this we introduce three pivotal concepts in probability theory: sample sapce $\mathscr{S}$, σ field $\mathscr{B}$ and probability measure $\mathbb{P}$.

2 Basic Concepts of Probability

Definition 2.1: Sample Space

The possible outcomes of a random experiment are called **basic outcomes** or **sample point**. the set of all basic outcomes constitutes the **sample space**, which is demoted by $\mathscr{S}$.

The basic outcomes are the basic building blocks for a sample space. They cannot be divided into more primitive outcomes. It is preferable that an element of a sample space does not represent two or more outcomes that are distinguishable in any way.

It is important to note that for a random experiment, one knows the set of all possible basic outcomes (i.e. the sample space), but does not know which outcome will arise before performing the random experiment. A sample space $\mathscr{S}$ can be countable or uncountable, finite or infinite. the distinction between a countable sample space and uncountable one dictates the ways where probabilities will be assigned.

Generally, we are interested in events constructed from basic outcomes, and investigate it. It includes the basic outcomes, but the contents is richer.

Definition 2.2: Event

A **event** A is a collection of basic outcomes from the sample space $\mathscr{S}$ that share certain common features or obey certain restriction. Mathematically, an event is equivalent to a set.

The event A is said to occur if the random experiment gives rise to any one of the constituent basic outcomes in $\mathcal{A}$. Obviously, we have the following relationship among the sample space, a basic outcome and an event:

$$\text{Basic outcomes} \subseteq \text{Event} \subseteq \text{Sample space}.$$

3 Review of Set Theory

Let A and B be two events (sets) in the sample space $\mathscr{S}$. Then we have the following definitions.

Definition 3.1

1. The **intersection** of A and B, denoted $A \cap B$, is the set of basic outcomes in $\mathscr{S}$ that belong to both A and B.
2. If $A \cap B = \varnothing$, we call A and B are **mutually exclusive**.
3. The **union** of A and B, denoted $A \cup B$, is the set of all basic outcomes in $\mathscr{S}$ that belong to either A or B.
4. Suppose $A_1, A_2, \cdots, A_n$ are n events in the sample space $\mathscr{S}$, where n is any positive integer. if $\bigcup_{i=1}^{n} A_i = \mathscr{S}$, then these n events are said to be **collectively exhaustive**.
5. The **complement** of A, denoted A^c, is the set of basic outcomes of' a random experiment that belong to $\mathscr{S}$ but not to A.
6. The **difference** of A and B, denoted as $A - B = A \cap B^c$, is the set of basic outcomes in $\mathscr{S}$ that belong to A but not to B.

It should be noted that, any pair of the basic outcomes in sample space $\mathscr{S}$ are mutually exclusive, because when a random experiment is performed, one and only one basic outcome will occur.

Obviously, any event A and its complement A^c are mutually exclusive and collectively exhaustive. that is $A \cap A^c = \varnothing$ and $A \cup A^c = \mathscr{S}$.

Theorem 3.1: Laws of Sets Operations

For any three events A, B, C defined on a sample space $\mathscr{S}$:

- *Complementation:*
$$(A^c)^c = A, \varnothing^c = \mathscr{S}, \mathscr{S}^c = \varnothing;$$
- *Commutativity of union and intersection:*
$$A \cup B = B \cup A$$
$$A \cap B = B \cap A;$$
- *Associativity of union and intersection:*
$$(A \cup B) \cup C = A \cup (B \cup C)$$
$$(A \cap B) \cap C = A \cap (B \cap C)$$
- *Distribution:*
$$(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$$
$$(A \cap B) \cup C = (A \cup C) \cap (B \cup C);$$

generally, for any $n \geq 1$:

$$B \bigcap \left(\bigcup_{i=1}^{n} A_i \right) = \bigcup_{i=1}^{n} (B \cap A_i)$$
$$B \bigcup \left(\bigcap_{i=1}^{n} A_i \right) = \bigcap_{i=1}^{n} (B \cup A_i);$$

- *De Morgan's Law:*

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

generally, for any $n \geq 1$:

$$\left(\bigcup_{i=1}^{n} A_i \right)^c = \bigcap_{i=1}^{n} A_i^c$$
$$\left(\bigcap_{i=1}^{n} A_i \right)^c = \bigcup_{i=1}^{n} A_i^c;$$

Proof. Here we prove the De Morgan's Law from the basic outcome perspective (which is a significant proof method in set theory):

If $x \in (A \cup B)^c$, then $x \in A^c$ and $x \in B^c$. It follows that $x \in A^c \cap B^c$.

Next, if $x \in A^c \cap B^c$, then $x \in A^c$ and $x \in B^c$. It follows that x does not belong to A and B (i.e. $A \cup B$). Therefore, $x \in (A \cup B)^c$ by the definition of the complement.

Now, we prove the second law by logical deduction (other important proof method):

we have $(A \cup B)^c = A^c \cap B^c$, therefore $((A \cup B)^c)^c = (A^c \cap B^c)^c = A \cup B$, let $A = A^c, B = B^c$, then $(A \cap B)^c = A^c \cup B^c$. The proof is completed. □

Example 3.1. Let the set of events $\{A_i, i = 1, \cdots, n\}$ be mutually exclusive and collectively exhaustive, and let A be an event in $\mathscr{S}$.

1. Are $A_i \cap A$ mutually exclusive ?
2. Is $\bigcup_{i=1}^{n} A \cap A_i$ equal to A ?

Solution. There are two ways to prove the first question:

By the basic outcome perspective: if $x \in A_i \cap A$, then $x \in A_i$.because $A_i \cap A_j = \varnothing$, we have $x \notin A_j$ that follows $x \notin A_j \cap A$. Therefore there is nothing in the intersection of $A_i \cap A$ and $A_j \cap A$, i.e. $(A_i \cap A) \cap (A_j \cap A) = \varnothing$.

By logical deduction:

$$\begin{aligned}(A_i \cap A) \cap (A_j \cap A) &= (A_i \cap A) \cap (A \cap A_j) = (A \cap A_j) \cap (A_i \cap A) \\ &= ((A \cap A_j) \cap A_i) \cap A = (A \cap (A_j \cap A_i)) \cap A \\ &= (A \cap \varnothing) \cap A = \varnothing \cap A = \varnothing.\end{aligned}$$

The second question:

$$\bigcup_{i=1}^{n} A \cap A_i = A \bigcap \left(\bigcup_{i=1}^{n} A_i \right) = A \cap \mathscr{S} = A.$$

□

4 Fundamental Probability Laws

We will assign a probability to an event A in $\mathscr{S}$. The Probability function is a function or a mapping from an event to a real number. More precisely, We are interested in assigning probability to *events, complement of events, unions and intersections of events*. Hence, we want our collection of events to include these combinations of events. Such a collection of events is called a σ field of subsets of the sample space $\mathscr{S}$, which will constitute the domain of the probability function.

Definition 4.1: σ Algebra

A σ **algebra**, denoted by $\mathscr{B}$, is a collection of subsets (events) of $\mathscr{S}$ that satisfy the following properties:

1. $\varnothing \in \mathscr{B}$;
2. if $A \in \mathscr{B}$, then $A^c \in \mathscr{B}$ (i.e. $\mathscr{B}$ is closed under complement);
3. if $A_1, A_2, \cdots \in \mathscr{B}$, then $\bigcup_{i=1}^{\infty} A_i \in \mathscr{B}$ (i.e. $\mathscr{B}$ is closed under countable union).

A σ algebra is also called a σ field. it is a collection of events in $\mathscr{S}$ that satisfy certain properties and the domain of a probability function on which we will assign a probability to any event.

It is important to note that a σ field is a collection of subsets in $\mathscr{S}$, but itself is not a subset of $\mathscr{S}$. One sample space $\mathscr{S}$ can produce multiple different sigma algebras $\mathscr{B}$. For any sample space, the smallest sigma algebra generated is $\mathscr{B} = \{\varnothing, \mathscr{S}\}$.The sample space $\mathscr{S}$ is only an element of a σ field. In probability theory, the event space is a σ field. The pair $(\mathscr{S}, \mathscr{B})$ is called a ***measurable space***.

Definition 4.2: Probability Function

Suppose a random experiment has a sample space $\mathscr{S}$, and a associated σ field $\mathscr{B}$. A **probability function** $\mathbb{P} : \mathscr{B} \mapsto [0, 1]$ is defined as a mapping that satisfies the following properties:

1. $0 \leq \mathbb{P} \leq 1$ for any event $A \in \mathscr{B}$;
2. $\mathbb{P}(\mathscr{S}) = 1$;
3. if $A_1, A_2, \cdots \in \mathscr{B}$ are mutually exclusive, then $\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$.

A Probability function tells how the probability of occurrence is distributed over the set of events, $\mathscr{B}$. In this sense we speak of a distribution of probability. Any function $f(\cdot)$ that satisfies the axioms of the probability is called a probability function. For a measurable space $(\mathscr{S}, \mathscr{B})$, many different probability functions can be defined. The goal of economics and statistics is to find a probability function that most accurately describes the underlying economic process. This probability function is usually called the true probability function or true probability distribution model.

So far we have defined the discussed concepts of sample space $\mathscr{S}$, σ field $\mathscr{B}$ and probability function $\mathbb{P}$ respectively. They together constitute a so-called probability space.

Definition 4.3: Probability Space

A **probability space** is a triple $(\mathscr{S}, \mathscr{B}, \mathbb{P})$ where:

1. $\mathscr{S}$ is the sample space corresponding to the outcomes of the underlying random experiment;
2. $\mathscr{B}$ is the σ field of subsets of $\mathscr{S}$. These subsets are called events;
3. $\mathbb{P} : \mathscr{B} \mapsto [0, 1]$ is a probability measure.

A probability space $(\mathscr{S}, \mathscr{B}, \mathbb{P})$ completely describes a random experiment associated with sample space $\mathscr{S}$. Because the probability function $\mathbb{P}(\cdot)$ is defined on $\mathscr{B}$, the collection of the events or sets, it is also called a set function. we now discuss some of its properties.

Theorem 4.1

If $\varnothing$ denotes the empty set, then $\mathbb{P}(\varnothing) = 0$.

Proof. Given $\mathscr{S} = \mathscr{S} \cup \varnothing$, and $\mathscr{S}$ and $\varnothing$ is mutually exclusive, we have $\mathbb{P}(\mathscr{P}) = \mathbb{P}(\mathscr{S} \cup \varnothing) = \mathbb{P}(\mathscr{S}) + \mathbb{P}(\varnothing) = 1$. It follows that $\mathbb{P}(\varnothing) = 0$. □

Intuitively, this theorem means that it is unlikely that nothing occurs when a random experiment is implemented. In other words, something always occurs when a random experiment is implemented.

While $\mathbb{P}(\varnothing) = 0$, it is important to note that it does not necessarily follow from $\mathbb{P}(A) = 0$ that $A = \varnothing$. Correspondingly, we also can not get $A = \mathscr{S}$ from $\mathbb{P}(A) = 1$. It will be clearly seen when we introduce a so-called continuous random variable in next chapter where a continuous random variable taking a single value has probability zero.

Theorem 4.2

$\mathbb{P}(A) = 1 - \mathbb{P}(A^c)$.

Proof. Obviously, A and A^c are mutually exclusive and collectively exhaustive, then we have

$$\mathbb{P}(A \cup A^c) = \mathbb{P}(A) + \mathbb{P}(A^c) = \mathbb{P}(\mathscr{S}) = 1.$$

The desired result follows immediately. □

To appreciate how useful this theorem is in practice, we consider a simple example:

Example 4.1. Suppose X demotes the outcome of some random experiment. The following is the probability distribution of X, namely, the probability that X takes various values:

$$\mathbb{P}(X = i) = \frac{1}{2^i},\ i = 1, 2, \cdots$$

Find the probability that X is lager than 3.

Solution. The sample space $\mathscr{S} = \{1, 2, \cdots\}$, Let A be the event that $X > 3$, then

$$\begin{aligned}
\mathbb{P}(A) &= 1 - \mathbb{P}(A^c) \\
&= 1 - \mathbb{P}(X = 1) - \mathbb{P}(X = 2) - \mathbb{P}(X = 3) \\
&= 1 - \frac{1}{2} - \frac{1}{2^2} - \frac{1}{2^3} \\
&= \frac{1}{8}.
\end{aligned}$$

□

Theorem 4.3

If A and B are two events in $\mathscr{B}$, and $A \subseteq B$, then $\mathbb{P}(A) \leq \mathbb{P}(B)$.

Proof. Given $B = B \cap \mathscr{S}$ and $\mathscr{S} = A \cup A^c$, we have that

$$\begin{aligned} B &= B \cap \mathscr{S} = B \cap (A \cup A^c) \\ &= (B \cap A) \cup (B \cap A^c) = A \cup (B \cap A^c) \end{aligned}$$

Because A and $B \cap A^c$ are mutually exclusive, we have

$$\mathbb{P}(B) = \mathbb{P}(A \cup (B \cap A^c)) = \mathbb{P}(A) + \mathbb{P}(B \cap A^c) \geq \mathbb{P}(A).$$

□

Corollary 4.1. *For any event $A \in \mathscr{B}$ such that $\varnothing \subseteq A \subseteq \mathscr{S}$, $0 \leq \mathbb{P}(A) \leq 1$.*

Theorem 4.4

For any two events A and B in $\mathscr{B}$,

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

Proof. Since $A \cup B = A \cup (B \cap A^c)$, and A and $B \cap A^c$ are mutually exclusive, we have

$$\mathbb{P}(A \cup B) = \mathbb{P}(A \cup (B \cap A^c)) = \mathbb{P}(A) + \mathbb{P}(B \cap A^c)$$

On the other hand, $B = B \cap \mathscr{S} = B \cap (A \cup A^c) = (B \cap A) \cup (B \cap A^c)$, and both $B \cap A$ and $B \cap A^c$ are mutually conclusive, we have

$$\mathbb{P}(B) = \mathbb{P}(B \cap A) + \mathbb{P}(B \cap A^c),$$

Namely

$$\mathbb{P}(B \cap A^c) = \mathbb{P}(B) - \mathbb{P}(B \cap A),$$

and

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

This delivers the desired result. □

Corollary 4.2 (Bonferroni's Inequality)**.** $\mathbb{P}(A \cup B) \geq \mathbb{P}(A) + \mathbb{P}(B) - 1$.

Theorem 4.5: Rule of Total Probability

If $A_1, A_2, \cdots \in \mathscr{B}$ are mutually exclusive and collectively exhaustive, and A is an event in $\mathscr{S}$, then

$$\mathbb{P}(A) = \sum_{i=1}^{\infty} \mathbb{P}(A \cap A_i).$$

Proof. Noting that $\mathscr{S} = \bigcup_{i=1}^{\infty} A_i$ and $A = A \cap \mathscr{S} = A \cap \bigcup_{i=1}^{\infty} A_i = \bigcup_{i=1}^{\infty} (A \cap A_i)$. This completes the proof. □

In certain sense, a set of mutually exclusive and collectively exhaustive events, $A_1, A_2, \cdots, A_n$, could be viewed as a set of complete ***orthogonal bases*** on which we can span any event A, where the intersection $A \cap A_i$ can be viewed as the projection of the event A onto base A_i.

Theorem 4.6: Subadditivity: Boole' Inequality

For any sequence of events $\{A_i \in \mathscr{B}, i = 1, 2, \cdots\}$,

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

Proof. Put $B = \bigcup_{i=2}^{\infty}$, then $\bigcup_{i=1}^{\infty} A_i = A_1 \cup B$.it follows that:

$$\begin{aligned}\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) &= \mathbb{P}\left(A_1 \cup B\right) \\ &= \mathbb{P}\left(A_1\right) + \mathbb{P}\left(B\right) - \mathbb{P}\left(A_1 \cap B\right) \\ &\leq \mathbb{P}\left(A_1\right) + \mathbb{P}\left(B\right).\end{aligned}$$

Put $C = \bigcup_{i=3}^{\infty} A_i$, then

$$\mathbb{P}\left(B\right) = \mathbb{P}\left(A_2 \cup C\right) \leq \mathbb{P}\left(A_2\right) + \mathbb{P}\left(C\right).$$

It follows that

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) \leq \mathbb{P}\left(A_1\right) + \mathbb{P}\left(A_2\right) + \mathbb{P}\left(C\right).$$

Repeating this process, we have the desired result. □

5 Methods of Counting

Suppose event A includes k basic outcomes $A_1, \cdots, A_k$ (the basic outcomes are mutually exclusive events) in sample space $\mathscr{S}$. Then

$$\mathbb{P}\left(S\right) = \sum_{i=1}^{k} \mathbb{P}\left(A_1\right).$$

This is the basic formula to compute the probability of any event A. If in addition $\mathscr{S}$ consists of n equally likely basic outcomes $A_1, \cdots, A_n$, then

$$\mathbb{P}\left(A_i\right) = \frac{1}{n}, \text{ for all } i = 1, \cdots, n.$$

In such a scenario, suppose event A consists of k basic outcomes. Then

$$\mathbb{P}\left(A\right) = \frac{k}{n}.$$

Thus, in the cases where each basic outcome is equally likely to occur, the calculation of probability for event A boils down to the ***counting of the numbers of basic outcomes in event*** A ***and in sample space*** $\mathscr{S}$ ***respectively*** . More generally, we can see the importance of counting methods in calculating the probability of events

Theorem 5.1: Fundamental Theorem of Counting

If a random experiment consists of k separate tasks, the i-th of which can be done in n_i *ways,* $i = 1, 2, \cdots, k$, *then the entire job can be done in* $n_1 \times n_2 \times \cdots \times n_k$ *ways.*

Proof. We shall first prove it for $k = 2$, then by induction. The first task can be done in n_1 ways, and for each ways, there are n_2 ways to do the second task. Therefore, the total ways for doing the first and second jobs is $n_1 \times n_2$. This completes the proof. □

We now investigate a ***specific random experiment*** where each basic outcome has k tasks and each task can be done in n ways where $n \geq k$. Each basic outcome is equally likely to occur, and given the fundamental theorem of counting, the total number of the basic outcomes in sample space $\mathscr{S}$ is n^k. To calculate probability of event from the $\mathscr{S}$, we need count the total number of the basic outcomes in event. Therefore, we will introduce two important counting methods – permutation and combination.

5.1 Permutation

Theorem 5.2: Permutation Formula

There are $\frac{n!}{(n-x)!}$ ways which is denoted by P_n^x to choose x objects from the n ones in ordered without replacement.
That is, for any basic outcome $x_i, x_j \in \mathscr{S}$ with n objects, if we consider the event that contains $x_i x_j$ is unequal to any event with $x_j x_i$, then we can crate the event with $P_n^x = \frac{n!}{(n-x)!}$ basic outcomes from the $\mathscr{S}$.

Proof. Firstly, we will choose the fist basic outcome from the $\mathscr{S}$. How many ways can we get it ? There are n objects available, so there are n ways.

Secondly, suppose we have chosen the first sample point, how many different ways can we select the second one ? Obviously, there exist $n-1$ objects remained and each of these can be used to the second choice. Therefore there are $n-1$ ways to get the second basic outcome.

Thus, there are $n \times (n-1)$ ways to obtain the first two basic outcomes.

By parity of reasoning, for the last choice, given that $x-1$ objects have been used to fill the first $x-1$ samples, there are $[n-(x-1)]$ objects left, so there are $[n-(x-1)]$ ways to opt.

To sum up, the total number of possible orderings of choosing x out of n basic outcomes is:

$$P_n^x = \frac{n!}{(n-x)!} = n \times (n-1) \times (n-2) \times \cdots \times [n-(x-1)].$$

□

Example 5.1. Suppose there are k students in a class, where $2 \leq k \leq 365$. What is the probability that at least two students have the same birthday?

Solution. First, how many possible ways in which the whole class could be born? This is a ***problem of ordering with replacement***:

$$365^k = \underbrace{365 \times 365 \times \cdots \times 365}_{k},$$

This is the total number of basic outcomes in the sample space $\mathscr{S}$.

Second, the event A where at least 2 students have the same birthday is complement to the event A^c where all k students have different birthday. Thus, the total number of the basic outcomes in event A^c (i.e. the number of ways that k students can have different birthday) is:

$$\frac{365!}{(365-k)!}$$

Therefore,

$$\begin{aligned}\mathbb{P}(A) &= 1 - \mathbb{P}(A^c) \\ &= 1 - \frac{365!}{365^k \times (365-k)!}\end{aligned}$$

The Matlab code is as follows:

```
x=0:1:100;
y=1-(factorial(sym(365)))./(factorial(sym(365-x)).*365.^x);
plot(x,y)
xlabel('the number of students of the class')
ylabel('probability')
```

□

5.2 Combination

We now consider the scenario of disordering without replacement. Suppose we are interested in the number of different ways of choosing x objects. Here, each object can be used at most once in each arrangement.

Theorem 5.3: Combination Formula

There are $\frac{n!}{(n-x)!x!}$ ways which is denoted by C_n^x to choose x objects from the n ones in disordered without replacement.

Proof. We consider the following basic formula (given the fundamental theorem of counting):

The total number of choosing x from n objects with ordering = The total number of choosing x from n objects without ordering $\times$ The number of ordering x objects.

The number of ordering x objects $= P_x^x = \frac{x!}{(x-x)!} = x!$, thus

$$C_n^x = \frac{P_n^x}{P_x^x} = \frac{n!}{(n-x)!x!}.$$

□

Example 5.2. Choose an integer randomly from 1 to 2000. How many possible ways to choose an integer that can be divided exactly neither by 6 nor by 8?

Solution. Define A = {the integer that can be divided exactly by 6}, and B = {the integer that can be divided exactly by 8}. Then by De Morgan's Law:

$$\begin{aligned}\mathbb{P}(A^c \cap B^c) &= \mathbb{P}((A \cup B)^c) \\ &= 1 - \mathbb{P}(A \cup B) \\ &= 1 - [\mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)].\end{aligned}$$

Because $333 < \frac{2000}{6} < 334$, then $\mathbb{P}(A) = \frac{333}{2000}$. Similarly, $\mathbb{P}(B) = \frac{250}{2000}$.

Moreover, an integer that can be divided exactly by both 6 and 8 is an integer that can be divided exactly by 24. Because $83 < \frac{2000}{24} < 84$, we have $\mathbb{P}(A \cap B) = \frac{83}{2000}$. It follows that:

$$\mathbb{P}(A^c \cap B^c) = 1 - [\mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)] = \frac{3}{4}.$$

The Matlab code is as follows:

```
k=50;
x=1:1:k;
[x,y]=meshgrid(x);
z=1-(floor(2000./x)./2000+floor(2000./y)./2000-floor(2000./lcm(x,y))./2000);
surf(x,y,z)
shading interp
xlabel('the first divisor')
ylabel('the second divisor')
zlabel('probability')
```

□

Example 5.3. We would like to choose r elements out of n elements. For the following cases, how many ways do we have?

1. ordered, without replacement;
2. disordered, without replacement;
3. ordered, with replacement;
4. disordered, with replacement.

Solution. Obviously, *1.* P_n^r; *2.* C_n^r; *3.* n^r. We now calculate the last one.

We choose n elements r times, suppose element i ($i = 1, 2, \cdots, n$) is chosen x_i times, then we have:

$$\sum_{i=1}^{n} x_i = x_1 + x_2 + \cdots + x_n = r.$$

Here x_i is non-negative integer. Let $y_i = x_i + 1$, we have:

$$\sum_{i=1}^{n} y_i = y_1 + y_2 + \cdots + y_n = r + n.$$

Here y_i is positive integer. Hence, we can apply the classic counting model that divide $n + r$ elements into n groups. We can choose $n - 1$ gaps from $n + r - 1$ gaps, thus the total number of the operation is C_{n+r-1}^{n-1}. This delivers the desired result. □

6 Conditional Probability

Economic events are generally related to each other. Because of the connection, the occurrence of event B may affect or contain the information about the probability that event A will occur. Thus if we have information about event B, then we can know better about the occurrence of event S. This can be described by the concept of conditional probability.

Intuitively, a sample space is a description of uncertainty we are faced with. The original sample space $\mathscr{S}$ of a random experiment describes the largest degree of uncertainty with which we view the experiment, When new information arrives, some uncertainty is eliminated. Specifically, when event B has occurred, the uncertainty is then reduced from $\mathscr{S}$ to B. In this case, we are in position to update the sample space based on new information. As a result, we want to be able to update the probability calculations as well. These updated probabilities are called conditional probabilities.

Definition 6.1: Conditional Probability

Let A and B be two events in probability space $(\mathscr{S}, \mathscr{B}, \mathbb{P})$. Then the **conditional probability** of event A given event B, denoted as $\mathbb{P}(A|B)$, is defined as

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

Here $\mathbb{P}(B) > 0$.

We provided that $\mathbb{P}(B) > 0$ because $\mathbb{P}(B) = 0$ implies that B is unlikely to happen, and conditioning on an unlikely event is meaningless from a practical point of view.

The conditional probability $\mathbb{P}(A|B)$ describes how to use the information on event to predict the probability of event A. This is a predictive relationship between A and B. It is important to note that

it is not necessarily a ***causal relationship*** from B to A, even if the information of event B can be used to predict event A. Thus if we want to be able to characterize a causal relationship, we have to use economic theory outside the probability and statistics.

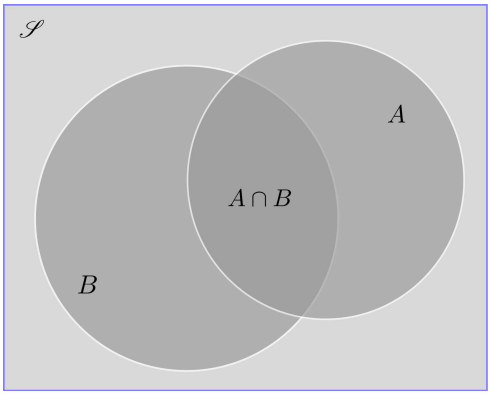

Figure 1: Venn diagram for conditional probability

The conditional probability $\mathbb{P}(A|B)$ can be represented in the Venn diagram. $\mathbb{P}(A|B)$ is the area occupied by event A within the area occupied by B (i.e. $A \cap B$) relative to the area occupied by B.

Specifically, when event B has occurred, the complement B^c will never occur. The uncertainty has been reduced from $\mathscr{S}$ to B. Thus, we will treat B as a new sample space when we consider $\mathbb{P}(A|B)$. Now, the triple $(\mathscr{S} \cap B, \mathscr{B} \cap B, \mathbb{P}(\cdot|B))$ is the new probability space associated with the conditional probability function $\mathbb{P}(A|B)$.

In particular, conditional probability satisfies all conditional counterparts of Theorem 4, 4, 4 and 4 hold.

Example 6.1. prove that

$$\mathbb{P}(A^c|B) = 1 - \mathbb{P}(A|B).$$

Solution. To show this, first, given the identity $(A^c \cap B) \cup (A \cap B) = B$ and the fact that the intersections $A^c \cap B$ and $A \cap B$ is mutually exclusive, we have

$$\mathbb{P}(B) = \mathbb{P}((A^c \cap B) \cup (A \cap B)) = \mathbb{P}(A^c \cap B) + \mathbb{P}(A \cap B),.$$

Thus

$$\mathbb{P}(A^c \cap B) = \mathbb{P}(B) - \mathbb{P}(A \cap B).$$

It follows that

$$\begin{aligned}\mathbb{P}(A^c|B) &= \frac{\mathbb{P}(A^c \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B) - \mathbb{P}(A \cap B)}{\mathbb{P}(B)} \\ &= 1 - \mathbb{P}(A|B).\end{aligned}$$

□

With the definition of conditional probability, we can state the following multiplication rules.

Lemma 6.1 (Multiplication Rules)**.** 1. *If* $\mathbb{P}(B) > 0$, *then* $\mathbb{P}(A \cap B) = \mathbb{P}(A|B) \cdot \mathbb{P}(B)$.

2. *If* $\mathbb{P}(A) > 0$, *then* $\mathbb{P}(A \cap B) = \mathbb{P}(B|A) \cdot \mathbb{P}(A)$.

Theorem 6.1

Suppose $\{A_i \in \mathscr{B},\ i = 1, 2, \cdots, n\}$ *is a sequence of n events. Then joint probability of the n events is*

$$\mathbb{P}\left(\bigcap_{i=1}^{n} A_i\right) = \prod_{i=1}^{n} \mathbb{P}\left(A_i \middle| \bigcap_{j=1}^{i-1} A_j\right),$$

with the convention that $\mathbb{P}\left(A_1 \mid \bigcap_{j=1}^{0} A_j\right) = \mathbb{P}(A_1)$.

Proof. We shall first prove it for $n = 3$, then by induction. Denote $B = A_1 \cap A_2$. Given lemma 6.1 we have

$$\begin{aligned}\mathbb{P}\left(\bigcap_{i=1}^{3} A_i\right) &= \mathbb{P}(B \cap A_3) = \mathbb{P}(A_3|B) \cdot \mathbb{P}(B) = \mathbb{P}(A_3|A_1A_2) \cdot \mathbb{P}(A_1A_2) \\ &= \mathbb{P}(A_3|A_1A_2) \cdot \mathbb{P}(A_2|A_1) \cdot \mathbb{P}(A_1) \\ &= \prod_{i=1}^{3} \mathbb{P}\left(A_i \middle| \bigcap_{j=1}^{2} A_j\right).\end{aligned}$$

□

These formula can be used to compute the joint probability of events by ***dividing them into a sequence of steps***.

Example 6.2 (Selecting Balls)**.** suppose 3 balls are to be selected, without replacement, from a box containing r red balls, b blue balls and k green balls. What is the probability that the first is red, the second is blue, and the third is green?

Solution. Define $A = \{\text{the first is red}\}$, $B = \{\text{the second is blue}\}$, $C = \{\text{the third is green}\}$. Then

$$\begin{aligned}\mathbb{P}(A) &= \frac{r}{r+b+k} \\ \mathbb{P}(B|A) &= \frac{b}{r+B+k-1} \\ \mathbb{P}(C|A \cap B) &= \frac{k}{r+b+k-2}\end{aligned}$$

It follows that

$$\begin{aligned}\mathbb{P}(A \cap B \cap C) &= \mathbb{P}(C|A \cap B) \cdot \mathbb{P}(B|A) \cdot \mathbb{P}(A) \\ &= \frac{rbk}{(r+b+k)(r+b+k-1)(r+b+k-2)}\end{aligned}$$

□

Theorem 6.2: Rule of Total Probability

Let $\{A_i\}_{i=1}^{\infty}$ *be a partition (i.e. mutually exclusive and collectively exhaustive) of sample space* $\mathscr{S}$*, with* $\mathbb{P}(A_i) > 0$ *for* $i \geq 1$*. Then for any event* $A in \mathscr{B}$*,*

$$\mathbb{P}(A) = \sum_{i=1}^{\infty} \mathbb{P}(A|A_i)\,\mathbb{P}(A_i).$$

Proof. Given Theorem 4 and Lemma 6.1, the desired result follows immediately. □

7 Bayes's Theorem

The knowledge that an event B has occurred can be used to revise or up date the prior probability that an event A will occur is the essence of Bayes's Theorem.

Theorem 7.1: Baye's Theorem

Suppose $A_1, \cdots, A_n$ are n mutually exclusive and collectively exhaustive events in the sample space $\mathscr{S}$, and the A is an event with $\mathbb{P}(A) > 0$. Then the conditional probability of A_i given A is

$$\mathbb{P}(A_i|A) = \frac{\mathbb{P}(A|A_i)}{\sum_{j=1}^{n} \mathbb{P}\left(A|A_j\right) \cdot \mathbb{P}\left(A_j\right)} \cdot \mathbb{P}(A_i).$$

Proof. By the conditional probability definition and multiplication rule, we have

$$\begin{aligned}\mathbb{P}(A_i|A) &= \frac{\mathbb{P}(A_i \cap A)}{\mathbb{P}(A)} \\ &= \frac{\mathbb{P}(A|A_i)}{\sum_{j=1}^{n} \mathbb{P}\left(A|A_j\right) \cdot \mathbb{P}\left(A_j\right)} \cdot \mathbb{P}(A_i).\end{aligned}$$

□

Here $\mathbb{P}(A_i)$ is called ***prior probability***, since it is the probability of A_i before the new information A arrives. the conditional probability $\mathbb{P}(A_i|A)$ is called ***posterior probability***, since it represents the revised assignment of probability of A_i after the new information that A has occurred is obtained. The bridge between the prior probability and the posterior probability, $\frac{\mathbb{P}(A|A_i)}{\sum_{j=1}^{n} \mathbb{P}\left(A|A_j\right) \cdot \mathbb{P}\left(A_j\right)}$, is called ***likelihood***.

The interest here is the update of the probability of event A_i when another event A has occurred. Bayes's Theorem shows how conditional probability of the form of $\mathbb{P}\left(A|A_j\right)$ may be combined with the prior probability $\mathbb{P}(A_i)$ to obtain the posterior probability $\mathbb{P}(A_i|A)$.

8 Independence

Definition 8.1: Independence

Events A and B are said to be statistically independent if

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\,\mathbb{P}(B).$$

Two events A and B are independent if the occurrence or non-occurrence of either one does not affect the probability of occurrence of the other, formally

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(A)\,\mathbb{P}(B)}{\mathbb{P}(B)} = \mathbb{P}(A).$$

Therefore, **the knowledge of B does not help in prediction** A, there is no causal relationship between them.

Independence is a probability notion to describe nonexistence of any kind of relation between two event. Generally, *mutually exclusive is also a kind of relation*, we now focus on the relation between the *independent* and *mutually exclusive* :

If A and B are independent with $\mathbb{P}(A) > 0$ and $\mathbb{P}(B) > 0$, then

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\,\mathbb{P}(B) > 0,$$

which means that $A \cap B \neq \varnothing$ and A and B are not mutually exclusive. On the other hand, if A and B are mutually exclusive events with positive probability, then

$$\mathbb{P}(A \cap B) = 0 \neq \mathbb{P}(A)\,\mathbb{P}(B),$$

which means that cannot be independent.

If A and B are independent with $\mathbb{P}(A) = 0$ or $\mathbb{P}(B) = 0$, then

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\,\mathbb{P}(B) = 0,$$

this implies that A and B could be mutually exclusive. On the other hand, if A and B are mutually exclusive events with zero probability, then

$$\mathbb{P}(A \cap B) = 0 = \mathbb{P}(A)\,\mathbb{P}(B),$$

they are independent.

The above discussion implies that the independent events with positive probability cannot be mutually exclusive, that is, independent events contain common basic outcomes and so can occur simultaneously. For instance, S&P 500 price index can increase when there is raining in Beijing.

Theorem 8.1

Let A and B are two independent events. Then

- *A and B^c are independent;*
- *A^c and B are independent;*
- *A^c and B^c are independent.*

Proof. Since $(A \cap B) cap (A \cap B^c) = \varnothing$ and $(A \cap B) \cup (A \cap B^c) = A$, we have that

$$\begin{aligned}\mathbb{P}(A \cap B^c) &= \mathbb{P}(A) - \mathbb{P}(A \cap B)\\ &= \mathbb{P}(A) - \mathbb{P}(A)\,\mathbb{P}(B)\\ &= \mathbb{P}(A)\,\mathbb{P}(B^c).\end{aligned}$$

The second statement could be proved by symmetry.

$$\begin{aligned}\mathbb{P}(A^c \cap B^c) &= \mathbb{P}((A \cup B)^c)\\ &= 1 - \mathbb{P}(A \cup B)\\ &= 1 - [\mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)]\\ &= [1 - \mathbb{P}(A)] - \mathbb{P}(B)\,[1 - \mathbb{P}(A)]\\ &= \mathbb{P}(A^c)\,\mathbb{P}(B^c).\end{aligned}$$

□

The above theorem could be understood intuitively: Suppose A and B are independent. Then A and B^c should be independent as well, otherwise one could be able to predict the probability of B^c using A, and thus predict the probability of B.

Definition 8.2: joint independence for events

The events $A_1, A_2, \cdots, A_n$ are joint independent, if

$$\mathbb{P}\left(\bigcap_{i\in I} A_i\right) = \prod_{i\in I} \mathbb{P}\left(A_i\right), \quad \text{for any finite } I \subset \{1, 2, \cdots, n\}.$$

We need $2^n - n - 1(\sum_{i=2}^{n} C_n^i)$ conditions to characterize independence among n events.

Obviously, by definition, joint independence implies pairwise independence. However, the converse is not true. Consider three events A_1, A_2, A_3, which are pairwise independent but not joint independence. It means that one uses A_2 or A_3 (not both) to predict A_1, then it is unhelpful; However, if one use A_2 and A_3 to predict, then A_1 is predictable.