Chapter 1 Foundation of Probability Theory

1.1 Review of Set Theory

Venn Diagram: Use circles (or other shapes) to denote sets (events). The interior of the circle represents the elements of the set, while the exterior represents elements which are not in the set.

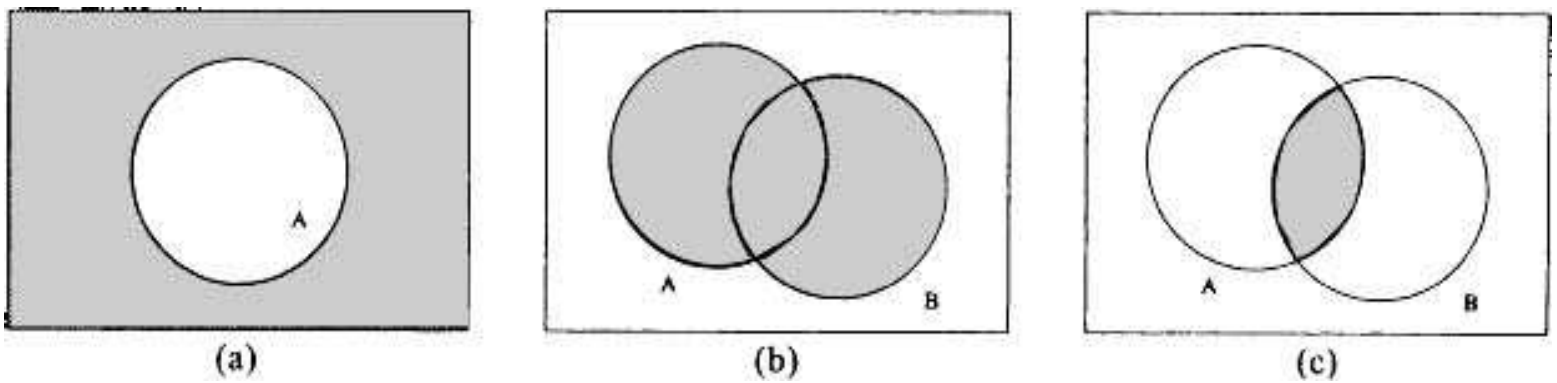

Figure 1: Venn Diagrams: (a) Complement , (b) Union, and (c) Intersection

1.1 Review of Set Theory

- **Definition: Containment.** The event A is *contained* in the event B, or B *contains* A, if every sample point of A is also a sample point of B. Whenever this is true, we will write $A \subseteq B$, or equivalently, $B \supseteq A$.

- **Definition: Equality.** Two events A and B are said to be *equal,* $A = B$, if $A \subseteq B$ and $B \subseteq A$.

- **Definition: Empty Set.** The set containing no elements is called the empty set and is denoted by $\emptyset$. The event corresponding to $\emptyset$ is called a *null* (*impossible*) event. An empty set $\emptyset$ is a subset of any set.

- **Definition: Intersection.** The *intersection* of A and B, denoted $A \cap B$, is the set of basic outcomes in S that belong to both A and B. The intersection occurs if and only if both events A and B occur.

- **Definition: Exclusiveness.** If A and B have no common basic outcomes, they are called *mutually exclusive* (or *disjoint*). Their intersection is empty set, *i.e.*, $A \cap B = \emptyset$.

- **Definition: Union.** The *union* of A and B, $A \cup B$, is the set of all basic outcomes in S that belong to either A or B. The union of A and B occurs if and only if either A or B (or both) occurs.

1.1 Review of Set Theory

- **Definition: Collective Exhaustiveness.** Suppose $A_1, A_2, \cdots, A_n$ are n events in the sample space S, where n is any positive integer. If $\cup_{i=1}^n A_i = S$, then these n events are said to be *collectively exhaustive*.

- **Definition: Complement.** The *complement* of A is the set of basic outcomes of a random experiment belonging to S but not to A, denoted as A^c.

- **Definition: Partition.** A class of events $\mathcal{H} = \{A_1, A_2, \cdots, A_n\}$ forms a partition of the sample space S if these events satisfy

(1) $A_i \cap A_j = \emptyset$ for all $i \neq j$ (mutually exclusive).

(2) $A_1 \cup A_2 \cup \cdots \cup A_n = S$ (collectively exhaustive).

- **Definition: Difference.** The *difference* of A and B, denoted by $A\backslash B$ or $A - B$, is the set of basic outcomes in S that belong to A but not to B, *i.e.*

$$A\backslash B = A \cap B^c.$$

The *symmetric difference*, $A \div B$, contains basic outcomes that belong to A or to B, but not to both of them.

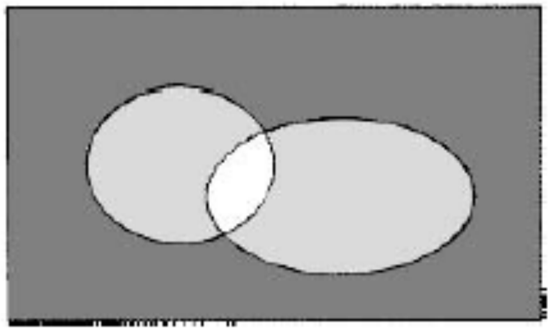

Figure 2: Venn Diagrams: Symmetric Difference.

1.1 Review of Set Theory : Laws of Sets Operations

Theorem: Laws of Sets Operations. For any three events A, B, C defined on a sample space S,

- Complementation

$$(A^c)^c = A, \qquad \emptyset^c = S, \qquad S^c = \emptyset$$

- Commutativity

$$A \cup B = B \cup A, \qquad A \cap B = B \cap A.$$

- Associativity

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

1.1 Review of Set Theory : Laws of Sets Operations

- Distributivity:

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C),$$

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

More Generally, for $n \geq 1$,

$$B \cap (\cup_{i=1}^{n} A_i) = \cup_{i=1}^{n} (B \cap A_i),$$

$$B \cup (\cap_{i=1}^{n} A_i) = \cap_{i=1}^{n} (B \cup A_i).$$

- De Morgan's Laws:

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

More Generally, for $n \geq 1$,

$$(\cup_{i=1}^{n} A_i)^c = \cap_{i=1}^{n} (A_i^c), \quad (\cap_{i=1}^{n} A_i)^c = \cup_{i=1}^{n} (A_i^c).$$

1.1 Review of Sets Theory : De Morgan's Laws

Use Venn diagrams to check De Morgan's Laws for the case of two events

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

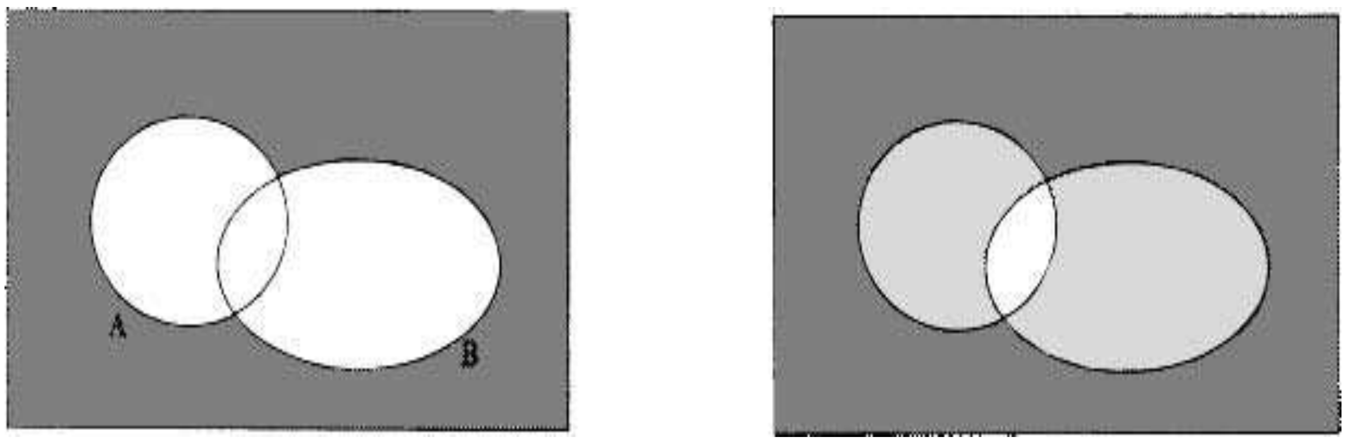

Figure 3: Validation of $(A \cup B)^c = A^c \cap B^c$. Left panel: $(A \cup B)^c$; right panel: $A^c \cap B^c$.

1.1 Review of Sets Theory : De Morgan's Laws

- How to prove the general case of De Morgan's Laws ($n > 2$)

$$(\cup_{i=1}^{n} A_i)^c = \cap_{i=1}^{n} (A_i^c).$$

- **Proof.**

$$\begin{aligned} & x \in (\cup_{i=1}^{n} A_i)^c \\ \Leftrightarrow & x \notin A_i \text{ for any } i = 1, \cdots, n \\ \Leftrightarrow & x \in A_i^c \text{ for any } i = 1, \cdots, n \\ \Leftrightarrow & x \in \cap_{i=1}^{n} (A_i^c). \end{aligned}$$

Hence, the equality holds.

1.2 Fundamental Probability Laws : Sigma Algebra

- **Definition: Sigma Algebra (or Sigma Field, Borel Field).** A *sigma algebra*, denoted by $\mathcal{B}$, is a **collection of subsets** of S with

  (1) $\emptyset \in \mathcal{B}$ (the empty set is contained in $\mathcal{B}$).

  (2) If $A \in \mathcal{B}$, then $A^c \in \mathcal{B}$ ($\mathcal{B}$ is closed under countable complement).

  (3) If $A_1, A_2, \cdots, \in \mathcal{B}$, then $\cup_{i=1}^{\infty} A_i \in \mathcal{B}$ ($\mathcal{B}$ is closed under countable unions).

- **Remarks:**
  - (1) and (2) imply that $S \in \mathcal{B}$.
  - (2) and (3) imply that $\cap_{i=1}^{\infty} A_i \in \mathcal{B}$.
  - For a given sample space S, we can construct many different σ-algebras.
  - A probability function is a function $P : \mathcal{B} \to [0, 1]$.

1.2 Fundamental Probability Laws : Sigma Algebra

- **Remarks:**
  - $\{\emptyset, S\}$ is a σ-algebra , usually called the *trivial* σ-algebra.
  - The collection of all possible subsets of S is a σ-algebra.
  - For any event A, $\{\emptyset, A, A^c, S\}$ is a σ-algebra.
  - Intersection of sigma algebras is also a σ-algebra.
  - The pair $(S, \mathcal{B})$ is called a measurable space.

- **Example:** Suppose the sample space $S = \{1, 2, 3\}$. Then, the set containing $\{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}, \emptyset$ is a σ-algebra.

1.2 Fundamental Probability Laws : Probability Function

- **Definition: Probability Function.** Suppose a random experiment has a sample space S and an associated σ-algebra $\mathcal{B}$. The probability function $P : \mathcal{B} \to [0, 1]$ is a mapping that satisfies the following properties:

(1) $0 \leq P(A) \leq 1$ for any event A in $\mathcal{B}$.

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

(3) If countable number of events $A_1, A_2, \cdots \in \mathcal{B}$ are mutually exclusive (pairwise disjoint), then $P(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$. (countable additivity.)

1.2 Fundamental Probability Laws : Interpretation of Probabil -ity

- Relative frequency interpretation: weather report says that there is a 30% chance for raining, it meas that under the same weather conditions it will rain 30% of the times.

- Subjective probability interpretation: subjective method, which, instead of thinking of probability as frequency, thinks of probability as a belief in the chance of an event occurring.

- Geometric representation of probability: Venn diagram, the probability of any event A in sample S can be viewed as equal to the area of event A in S, with the normalization that the total area of S is equal to unity.

1.2 Fundamental Probability Laws : Probability Space

- **Definition: Probability Space.** A *probability space* is a triple $(S, \mathcal{B}, P)$, where
  - S is the sample space corresponding to outcomes of the underlying random experiment.
  - $\mathcal{B}$ is an associated σ-algebra of S. Elements of $\mathcal{B}$ are subsets of S. These subsets are called events.
  - P is a probability measure (probability function).

Theorems

- $P(\Phi) = 0$.
- $P(A^c) = 1 - P(A)$.

Example. Suppose X denotes the outcome if some random experiment. The probability distribution for X is $P(X = i) = \frac{1}{2^i}, i = 1, 2, ...$ Find the probability that X is larger than 3.

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

Theorems

- If $A \subseteq B$, $P(A) \leq P(B)$.
  Proof. $P(B) = P(A) + P(A^c \cap B) \geq P(A)$.

- (Corollary) For any event $A \in \mathcal{B}$ such that $\emptyset \subseteq A \subseteq S$, $0 \leq P(A) \leq 1$.

- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.
  Proof. Since $A \cup B = A \cup (A^c \cap B)$, A and $A^c \cap B$ are mutually exclusive, so

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

On the other hand, $B = S \cap B = (A \cap B) \cup (A^c \cap B)$, so

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

Then, it's easy to finish the proof.

Example [Bonferroni's inequality]. Show $P(A \cup B) \geq P(A) + P(B) - 1$.

Proof. Since $A \cap B \subseteq S$, then $P(A \cap B) \leq P(S) = 1$. It follow from theorem that

$$\begin{aligned} P(A \cup B) &= P(A) + P(B) - P(A \cap B) \\ &\geq P(A) + P(B) - 1. \end{aligned}$$

Theorem [Rule of Total Probability]. If $A_1, A_2, \cdots$ are mutually exclusive and collectively exhaustive, then

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

Proof. Noting $S = \cup_{i=1}^{\infty} A_i$ and $A = A \cap S = A \cap (\cup_{i=1}^{\infty} A_i) = \cup_{i=1}^{\infty} (A \cap A_i)$.The result follows because $A \cap A_i$ and $A \cap A_j$ are disjoint for all $i \neq j$.

Subadditivity [Boole's inequality]: For any sequence of events A_i, $i = 1, 2, \cdots$,

$$P(\cup_{i=1}^{\infty} A_i) \leq \sum_{i=1}^{\infty} P(A_i).$$

Proof. Put $B = \cup_{i=2}^{\infty} A_i$. Then $\cup_{i=1}^{\infty} A_i = A_1 \cup B$. It follows that

$$\begin{aligned} P(\cup_{i=1}^{\infty} A_i) &= P(A_1 \cup B) \\ &= P(A_1) + P(B) - P(A_1 \cap B) \\ &\leq P(A_1) + P(B), \end{aligned}$$

where the inequality follows given $P(A_1 \cap B) \geq 0$. Again, put $C = \cup_{i=3}^{\infty} A_i$. Then $P(B) = P(A_2 \cup C) \leq P(A_2) + P(C)$. It follows that

$$P(\cup_{i=1}^{\infty} A_i) \leq P(A_1) + P(A_2) + P(C).$$

Repeating this process, we have $P(\cup_{i=1}^{\infty} A_i) \leq \sum_{i=1}^{\infty} P(A_i)$.

1.3 Methods of Counting

- For the so-called *classical* or *logical interpretation* of probability, we will assume that the sample space S contains a finite number N of outcomes and all of these outcomes are equally probable.

- The associated sigma-algebra is the collection of all possible subsets of S.

- For every event A,

$$P(A) = \frac{\text{number of outcomes in } A}{N}.$$

- How to determine the number of total outcomes in the space S and in various events in S?

1.3 Methods of Counting

- We consider two important counting methods : *permutation* and *combination*.

- **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, $i = 1, \cdots, k$, then the entire job can be done in $n_1 \times n_2 \times \cdots \times n_k$ ways.

- **Example: Permutation.** Suppose we will choose two letters from four letters $\{A, B, C, D\}$ in different orders, with each letter being used at most once each time. How many possible orders could we obtain?

  There are 12 ways: $\{AB, BA, AC, CA, AD, DA, BC, CB, BD, DB, CD, DC\}$. The word "*ordered*" means that AB and BA are distinct outcomes.

1.3 Methods of Counting : Permutation

Permutations

- **Problem:** Suppose that there are k boxes arranged in row and there are n objects, where $k \leq n$. We are going to choose k from the n objects to fill in the k boxes. How many possible different **ordered sequences** could you obtain?
  - First, one object is selected to fill in box 1, there are n ways.
  - A second object is selected from the remaining $n-1$ objects. Therefore, there are $n-1$ ways to fill box 2.
  - $\vdots$
  - The last box (box k), there are $n-(k-1)$ ways to fill it.

1.3 Methods of Counting : Permutation

- The total number of different ways to fill box $1, 2, \cdots, k$ is

$$n(n-1)\cdots(n-(k-1)) = \frac{n!}{(n-k)!}.$$

- The experiment is equal to selecting k objects out of the n objects first, then arrange the selected k objects in a sequence.

- Each different arrangement of the sequence is called a *permutation*.

- The number of permutations of choosing k out of n, denoted by P_n^k, is

$$P_n^k = \frac{n!}{(n-k)!}.$$

- **Convention:** $0! = 1$.

1.3 Methods of Counting : Permutation

- **Example : The Birthday Problem.** What is the probability that at least two people in a group of k people ($2 < k \leq 365$) will have the same birthday?
  - Let $S = \{(x_1, x_2, \cdots, x_k)\}$, x_i represents the birthday of person i. How many outcomes in S? 365^k.
  - How many ways the k people can have different birthdays? P_{365}^k.
  - The probability that all k people will have different birthday is $P_{365}^k/365^k$.
  - The probability that at least two people will have the same birthday is $p = 1 - P_{365}^k/365^k$.
  - $k = 10, p = 0.1169482$; $k = 20, p = 0.4114384$, $k = 30, p = 0.7063162$, $k = 40, p = 0.8912318$, $k = 50, p = 0.9703736$.

1.3 Methods of Counting : Combination

Combinations

- Choosing a subset of k elements from a set of n distinct elements.
- The order of the elements is irrelevant. For example, the subsets $\{a, b\}$ and $\{b, a\}$ are identical.
- Each subset is called a *combination*.
- The number of combinations of choosing k out of n is denoted by C_n^k. We have

$$\begin{aligned} C_n^k &= \frac{\text{the number of choosing } k \text{ out of } n \text{ with ordering}}{\text{the number of ordering } k \text{ elements}} \\ &= P_n^k / k! \\ &= n!/k!(n-k)!. \end{aligned}$$

- **Example: Combination.** Suppose we will choose two letters from four letters $\{A, B, C, D\}$. Each letter is used at most once in each arrangement but now we are not concerned with their ordering. How many possible pairs could we have?

  There are six pairs: $\{A, B\}, \{A, C\}, \{A, D\}, \{B, C\}, \{B, D\}, \{C, D\}$.

- **Example:** A class contains 15 boys and 30 girls, and 10 students are to be selected at random for a special assignment. What is the probability that exactly 3 boys will be selected?

  - The number of combinations of 10 students out of 45 students is C_{45}^{10}.
  - The number of combinations of 3 boys out of 15 boys is C_{15}^{3}.
  - The number of combinations of 7 girls out of 30 girls is C_{30}^{7}.
  - Thus, $p = C_{15}^{3} C_{30}^{7} / C_{45}^{10}$.

1.3 Methods of Counting : Combination

- C_n^k is also denoted by $\begin{pmatrix} n \\ k \end{pmatrix}$. This is also called a *binomial coefficient* because of its appearance in the *binomial theorem*

$$(x+y)^n = \sum_{k=0}^{n} \begin{pmatrix} n \\ k \end{pmatrix} x^k y^{n-k}.$$

- **Properties of Combinations.**

1. $\begin{pmatrix} n \\ k \end{pmatrix} = \begin{pmatrix} n \\ n-k \end{pmatrix}$;

2. $\begin{pmatrix} n \\ 1 \end{pmatrix} = n$;

3. $\begin{pmatrix} n \\ k \end{pmatrix} = \frac{P_n^k}{k!}$.

- **Example B(n, p).** Suppose we throw a fair coin n times independently. What is the probability that x heads will show up?

  Solution. There are $2 \times 2 \times \cdots \times 2 = 2^n$ outcomes in total. And there are a total of C_n^x different ways to obtain x heads. So

$$P(\text{exactly } x \text{ heads}) = C_n^x / 2^n.$$

- **Example (Draw balls).** We want to choose r elements out of n elements. For the following cases, how many ways do we have ?

  - *ordered, without replacement.* $P_n^r = \frac{n!}{(n-r)!}$;
  - *unordered, without replacement.* $C_n^r = P_n^r / r! = \frac{n!}{r!(n-r)!}$;
  - *ordered, with replacement.* n^r;
  - *unordered, with replacement.* C_{r+n-1}^r.

1.4 Conditional Probability

- Different 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.

- **Example: Financial Contagion.** A large drop of the price in one market can cause a large drop of the price in another market, given the speculations and reactions of market participants.

1.4 Conditional Probability

- **Definition: Conditional Probability.** Let A and B be two events in $(S, \mathcal{B}, P)$. Then the conditional probability of event A given event B, denoted as $P(A \mid B)$, is defined as

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

provide that $P(B) > 0$.

- **Remarks:**
  - $P(\cdot \mid B)$ is a new probability function defined on $(S, \mathcal{B})$.
  - We can treat B as a new sample space when we consider $P(A \mid B)$.
  - $(B, \mathcal{B} \cap B, P(\cdot \mid B))$ is a new probability space induced by event B.

- **Properties of Conditional Probability:**
  - $P(A) = P(A \mid S)$.
  - $P(A^c \mid B) = 1 - P(A \mid B)$.
  - *Multiplication Rules* $P(B) > 0$, $P(A) > 0$

$$P(A \cap B) = P(B)P(A \mid B) = P(A)P(B \mid A).$$

- **Example:** Suppose two balls are to be selected, without replacement, from a box containing r red balls and b blue balls. What is the probability that the first is red and the second is blue?

  ANS: Let $A = \{\text{the first ball is red}\}$, $B = \{\text{the second ball is blue}\}$. Then

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

1.4 Conditional Probability

- **Theorem: Chain Rule.** For any events $A_1, A_2, \cdots, A_n$, we have

$$P(\cap_{i=1}^{n} A_i) = \prod_{i=1}^{n} P(A_i | \cap_{j=1}^{i-1} A_j),$$

with the convention that $P(A_1 | \cap_{j=1}^{0} A_j) = P(A_1)$.

- **Example. [Computation of Joint Probabilities].** For $n = 3$, we have

$$P(A_1 \cap A_2 \cap A_3) = P(A_3 | A_2 \cap A_1) P(A_2 | A_1) P(A_1).$$

Remarks. In time series analysis, where i is an index for time, the partition that the event A_i is conditional on $\cap_{j=1}^{i-1} A_j$ has a nice interpretation: A_i is conditional on the past information available at time $i - 1$.

- **Theorem: Rule of Total Probability.** Let $\{A_i\}_{i=1}^{\infty}$ be a partition (i.e., mutually exclusive and collectively exhaustive) of sample space S, with $P(A_i) > 0$ for $i \geq 1$. For any event A in S,

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

  Intuition If event B can be partitioned as a set of mutually exclusive subevents, then the probability of event B is equal to the sum of probabilities of this set of mutually exclusive subevents contained in B. It is also called the *rule of elimination.*

- **Example:** Suppose B_1, B_2, and B_3 are mutually exclusive. If $P(B_i) = 1/3$ and $P(A \mid B_i) = i/6$ for $i = 1, 2, 3$. What is $P(A)$? (Hint: B_1, B_2, B_3 are also collectively exhaustive.)

1.5 Bayesian Analysis

- **Bayes' Theorem:** Suppose A and B are two events and $P(B) > 0$. Then

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{P(B \mid A)P(A)}{P(B \mid A)P(A) + P(B \mid A^c)P(A^c)}.$$

- **Remarks:**
  - We consider $P(A)$ as the prior probability about the event A.
  - $P(A \mid B)$ is posterior probability given that B has occurred.
  - Verbally, the posterior probability of event A is proportional to the probability of the same evidence B after A has occurred times the prior probability of A.
  - A "backward" or "inverse" sort of reasoning, "from effect to cause".

Alternative Statement of Bayes' Theorem: Suppose $A_1, \dots, A_n$ are n mutually exclusive and collectively exhaustive events in the sample space S, and A is an event with $P(A) > 0$. Then the conditional probability of A_i given A is

$$\begin{aligned} P(A_i \mid A) &= \frac{P(A \cap A_i)}{P(A)} \\ &= \frac{P(A \mid A_i)P(A_i)}{\sum_{j=1}^{n} P(A \mid A_j)P(A_j)}, \qquad i = 1, \dots, n. \end{aligned}$$

Proof.

$$\begin{aligned} P(A_i \mid A) &= P(A_i \cap A)/P(A) \\ &= P(A \mid A_i)P(A_i)/P(A) \\ &= \frac{P(A \mid A_i)P(A_i)}{\sum_{j=1}^{n} P(A \cap A_j)} \\ &= \frac{P(A \mid A_i)P(A_i)}{\sum_{j=1}^{n} P(A \mid A_j)P(A_j)}. \end{aligned}$$

- **Example: Auto-insurance.** Suppose an insurance company has three types of customers : high risk, medium risk and low risk. From the company's consumer database, it is known that 25% of its customers are high risk, 25% are medium risk, and 50% are low risk. Also, the database shows that the probability that a customer has at least one speeding ticket in one year is 0.25 for high risk, 0.16 for medium risk, and 0.10 for low risk.

  What is the probability that a new customer is high risk, given that he has had one speeding ticket this year?

- **Solution:** Let H, M, L are the events that the customer is a high risk, medium risk and low risk customer. Let A be the event that the customer has received a speeding ticket this year. Then
$$P(H \mid A) = \frac{P(A \mid H)P(H)}{P(A \mid H)P(H) + P(A \mid M)P(M) + P(A \mid L)P(L)}.$$
Given $P(H) = 0.25$, $P(M) = 0.25$, $P(L) = 0.50$, $P(A \mid H) = 0.25$, $P(A \mid M) = 0.16$, $P(A \mid L) = 0.10$, we have
$$P(H \mid A) = 0.410.$$

- Without the speeding ticket information reported by the new customer, the auto-insurance company, based on its customer database, only has a prior probability $P(H) = 0.25$ for the new customer. With the new information (A), the auto-insurance company has an updated probability $P(H \mid A) = 0.41$ for the new customer.

1.5 Bayesian Analysis

Example: A publisher sends a sample statistics textbook to 80% of all statistics professors in the U.S. schools. 30% of the professors who receive this sample textbook adopt the book, as do 10% of the professors who do not received the sample book. What is the probability that a professor who adopts the book has received a sample book?

Solution: Define event $A = \{$a professor receives a sample copy$\}$. Then,

$$P(A) = 0.8,\ P(A^c) = 1 - 0.8 = 0.2.$$

Define $B = \{$the professor adopts the textbook$\}$. Then

$$P(B \mid A) = 0.3,\ P(B \mid A^c) = 0.1.$$

It follows from Bayes' theorem that

$$\begin{aligned} P(A \mid B) &= P(A \cap B)/P(B) \\ &= \frac{P(B \mid A)P(A)}{P(B \mid A)P(A) + P(B \mid A^c)P(A^c)} \\ &= \frac{0.3 \cdot 0.8}{0.3 \cdot 0.8 + 0.1 \cdot 0.2} \\ &= 0.923. \end{aligned}$$

1.6 Independence

- If two events A and B are unrelated, we expect that the information of B is irrelevant to predicting $P(A)$. In other words, we expect that $P(A \mid B) = P(A)$.

- **Definition: Independence.** Events A and B are said to be statistically independent if $P(A \cap B) = P(A)P(B)$.

- **Remarks.** Independent, statistically independent, stochastically independent, independent in a probability. It is a statistical notion to describe nonexistence of any kind of relationship between two events.

- **Remarks:**
  - By this definition,

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

    Similarly, we have $P(B|A) = P(B)$. Therefore, **the knowledge of B does not help in predicting A**.

- **Example.** Let $A = \{\text{Raining in Xiamen}\}$, $B = \{\text{Standard \& Poor 500 price index going up}\}$. These two events are likely to be independent.

- **Example: Phillips Curve.** Let $A = \{\text{Inflation rate increases}\}$; $B = \{\text{Unemployment decreases}\}$. These two events are likely to be dependent of each other.

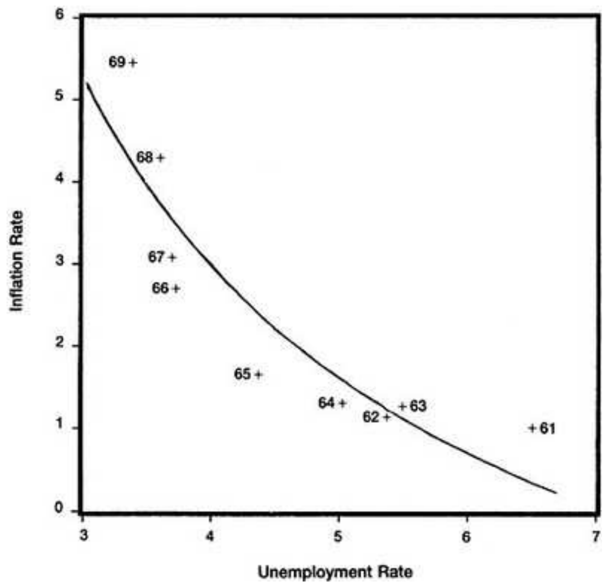

Figure 4: Phillips Curve.

1.6 Independence

- **Example: Random Walk Hypothesis (Fama 1970).** If a stock market is fully efficient, then the stock price P_t will follow a random walk; that is, $P_t = P_{t-1} + X_t$, where the stock price change $\{X_t = P_t - P_{t-1}\}$ is independent across different periods.

- **Example: Geometric Random Walk Hypothesis.** The stock price $\{P_t\}$ is called a geometric random walk if $X_t = \ln P_t - \ln P_{t-1}$ is independent across different time periods. Note that $X_t \approx \frac{P_t - P_{t-1}}{P_{t-1}}$ approximates the relative stock price change.

- **Remarks:** If $\{X_t\}$ is serially independent across different time periods, then a future stock price change X_t is not predictable using the historical stock price information. In such a case, we call the stock market is informationally efficient.

1.6 Independence

- **Example:** Can two independent events A and B be mutually exclusive? Can two mutually exclusive events A and B be independent?

- **Solution:**

Case A: If A and B are independent with $P(A) > 0$ and $P(B) > 0$, then

$$P(A \cap B) = P(A)P(B) > 0.$$

Therefore, if A and B are independent, they cannot be mutually exclusive. If A and B are mutually exclusive (so $P(A \cap B) = 0$), they can not be independent.

Case B: Suppose $P(A) = 0$, or $P(B) = 0$. If A and B are independent, then

$$P(A \cap B) = P(A)P(B) = 0.$$

This implies that A and B could be mutually exclusive. On the other hand, if A and B are mutually exclusive, they are independent.

1.6 Independence

- **Theorem:** Let A and B be two independent events. Then (a) A and B^c; (b) A^c and B; (c) A^c and B^c are all independent.

- **Proof.** (a)

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

  A and B^c are independent.

- **Remark:** Intuitively, A and B^c should be independent. Because if not, we would be able to predict B^c from A, and thus predict B.

1.6 Independence

- **Definition: Independence Among Several Events.** k Events $A_1, \cdots, A_k$ are mutually independent if, for every possible subset $A_{i_1}, \cdots, A_{i_j}$ of j of those events ($j = 2, \cdots, k$),

$$P(A_{i_1} \cap \cdots \cap A_{i_j}) = P(A_{i_1}) \cdots P(A_{i_j}).$$

- **Remark:** We need to verify $2^k - 1 - k$ conditions. For example, three events A, B, and C are independent if

$$\begin{aligned}
P(A \cap B) &= P(A)P(B), \\
P(A \cap C) &= P(A)P(C), \\
P(B \cap C) &= P(B)P(C), \\
P(A \cap B \cap C) &= P(A)P(B)P(C).
\end{aligned}$$

- **Remark:** It is possible to find that three events are mutually (pairwise) independent but not jointly independent. It is also possible to find three events A, B, C that satisfy $P(A \cap B \cap C) = P(A)P(B)P(C)$ but not independent.

- **Example:** Suppose

$$S = \{aaa, bbb, ccc, abc, bca, cba, acb, bac, cab\}$$

and each basic outcome is equally likely to occur. Define

$$A_i = \{i\text{-th place in the triple is occupied by letter } a\}, i = 1, 2, 3.$$

For example, $A_1 = \{aaa, abc, acb\}$. Are A_1, A_2, A_3 mutually independent?

- **Solution:** It is easy to see that

$$P(A_1) = P(A_2) = P(A_3) = \frac{1}{3},$$

and

$$P(A_1 \cap A_2) = P(A_1 \cap A_3) = P(A_2 \cap A_3) = \frac{1}{9},$$

so that A_1, A_2 and A_3 are pairwise independent. However,

$$P(A_1 \cap A_2 \cap A_3) = \frac{1}{9} > P(A_1)P(A_2)P(A_3).$$

Therefore, A_1, A_2, A_3 are not mutually independent.

- **Conclusion of this Chapter**

Have a happy holiday!