Chapter 3
Conditional Probability and Independence

Lectures 8 -12
In this chapter, we introduce the concepts of conditional probability and independence.

Suppose we know that an event \( A \) already occurred. Then a natural question asked is “What is the effect of this on the probabilities of other events?” This leads to the notion of conditional probability.

**Definition 3.1** Let \((\Omega, \mathcal{F}, P)\) be a probability space and \( A \in \mathcal{F} \) be such that \( P(A) > 0 \). The conditional probability of an event \( B \in \mathcal{F} \) given \( A \) denoted by \( P(B | A) \) is defined as

\[
P(B | A) = \frac{P(AB)}{P(A)}.
\]

Define

\[
\mathcal{F}_A = \{BA \mid B \in \mathcal{F}\}.
\]

Then \( \mathcal{F}_A \) is a \( \sigma \)-field of subsets of \( A \) (see Exercise 3.2).

A collection \( \{A_1, A_2, \ldots, A_N\} \) of events is said to be a *partition of* \( \Omega \), if

(i) \( A_i \)'s are pairwise disjoint

(ii) \( \bigcup_{i=1}^N A_i = \Omega \).

Here \( N \) may be \( \infty \). If \( N < \infty \), then partition is said to be finite partition and if \( N = \infty \), it is called a countable partition.

**Theorem 3.0.8** Define \( P_A \) on \( \mathcal{F}_A \) as follows.

\[
P_A(B) = P(B | A), \ B \in \mathcal{F}_A.
\]

Then \( (A, \mathcal{F}_A, P_A) \) is a probability space.

Proof is an exercise, see Exercise 3.3.

**Theorem 3.0.9** (Law of total probability-discrete form) Let \((\Omega, \mathcal{F}, P)\) be a probability space and \( \{A_1, A_2, \ldots, A_n\} \subseteq \mathcal{F} \) be a partition of \( \Omega \) such that \( P(A_i) > 0 \) for all \( i \). Then for \( B \in \mathcal{F} \),

\[
P(B) = \sum_{i=1}^n P(B | A_i) P(A_i)
\]
Proof.

\[
\sum_{i=1}^{n} P(B|A_i)P(A_i) = \sum_{i=1}^{n} \frac{P(BA_i)}{P(A_i)}P(A_i) \\
= \sum_{i=1}^{n} P(B|A_i) = P(B(\cup_{i=1}^{n} A_i)) = P(B)
\]

**Theorem 3.0.10** (Bayes Theorem) Let \((\Omega, \mathcal{F}, P)\) be a probability space and \(A, B \in \mathcal{F}\) such that \(P(A), P(B) > 0 \) and \(P(A) < 1\). Then

\[
P(A|B) = \frac{P(B|A)P(A)}{P(B|A) + P(B|A^c)}.
\]

Proof.

\[
P(A|B) = \frac{P(BA)P(A)}{P(A)P(B)} = \frac{P(B|A)P(A)}{P(B)}
\]

Now use Law of total probability to complete the proof.

**Definition 3.2** (Independence) Two events \(A, B\) are said to be independent if

\[P(AB) = P(A)P(B),\]

**Remark 3.0.3** If \(P(A) > 0\), then \(A\) and \(B\) are independent iff \(P(B|A) = P(A)\). This confirms the intuition behind the notion of independence, "the occurrence one event doesn't have any effect on the occurrence of the other".

**Example 3.0.17** Define the probability space \((\Omega, \mathcal{F}, P)\) as follows.

\[
\Omega = \{HH, HT, TH, TT\}, \mathcal{F} = \mathcal{P}(\Omega)
\]

and

\[P(\{\omega\}) = \frac{1}{4}, \omega \in \Omega.\]

Consider the events \(A = \{HH, HT\}, B = \{HH, TH\}\) and \(C = \{HT, TT\}\). Then \(A\) and \(B\) are independent and \(B\) and \(C\) are dependent.

**Example 3.0.18** Consider the probability space defined by

\[
\Omega = (0, 1], \mathcal{F} = B(0, 1]
\]
and $P$ given by (1.0.6). Consider the events $A = (0, \frac{1}{2}]$, $B = [\frac{1}{4}, \frac{3}{4}]$, $C = [\frac{1}{4}, 1]$. Then $A$, $B$ are independent and $A$, $C$ are dependent.

The notion of independence defined above can be extended to independence of three or more events in the following manner.

**Definition 3.3** (Independence of three events). The events $A$, $B$, $C$ are independent (mutually) if

(i) $A$, $B$; $B$, $C$ and $C$, $A$ are independent and

(ii) $P(ABC) = P(A)P(B)P(C)$.

If the events $A$, $B$, $C$ satisfies only (i), then $A$, $B$, $C$ are said to be pairwise independent.

**Example 3.0.19** Define $\Omega = \{1, 2, 3, 4\}$, $\mathcal{F} = \mathcal{P}(\Omega)$ and

$$P(\{i\}) = \frac{1}{4}, \ i = 1, 2, 3, 4.$$  

Consider the events $A = \{1, 2\}$, $B = \{1, 3\}$, $C = \{1, 4\}$. Then $A$, $B$, $C$ are pairwise independent but not independent.

**Example 3.0.20** Define $(\Omega, \mathcal{F}, P)$ as follows. $\Omega = \{1, 2, 3, 4, 5, 6, 7, 8\}$, $\mathcal{F} = \mathcal{P}(\Omega)$ and

$$P(\{i\}) = \frac{1}{8}, \ i = 1, \ldots, 8.$$  

Let $A = \{1, 2, 3, 4\}$, $B = \{3, 4, 5, 6\}$, $C = \{1, 3, 5, 7\}$. Then $A$, $B$, $C$ are independent.

**Definition 3.4** The events $\{A_1, A_2, \ldots, A_n\} \subseteq \mathcal{F}$ are said to be independent if for any distinct $A_{i_1}, \ldots, A_{i_m}, m \geq 2$ from $\{A_1, A_2, \ldots, A_n\}$

$$P(A_{i_1} \ldots A_{i_m}) = P(A_{i_1}) \ldots P(A_{i_m}).$$

Using the notion of independence of events, one can define the independence of $\sigma$-fields as follows.
Definition 3.5 Two \( \sigma \)-fields \( \mathcal{F}_1, \mathcal{F}_2 \) of subsets of \( \Omega \) are said to be independent if for any \( A \in \mathcal{F}_1, \ B \in \mathcal{F}_2 \)

\[
P(AB) = P(A)P(B).
\]

Definition 3.6 The \( \sigma \)-fields \( \mathcal{F}_1, \mathcal{F}_2, \ldots, \mathcal{F}_n \) of subsets of \( \Omega \) are said to be independent if \( A_1, A_2, \ldots, A_n \) are independent whenever \( A_i \in \mathcal{F}_i, \ i = 1, 2, \ldots, n. \)

One can go a step further to define independence of any family of \( \sigma \)-fields as follows.

Definition 3.7 A family of \( \sigma \)-fields \( \{ \mathcal{F}_i \mid i \in I \} \), where \( I \) is an index set, are independent if for any finite subset \( \{ \alpha_1, \cdots \alpha_n \} \subseteq I \), the \( \sigma \)-fields \( \mathcal{F}_{\alpha_1}, \cdots, \mathcal{F}_{\alpha_n} \) are independent.

Finally one can introduce the notion of independence of random variables through the corresponding \( \sigma \)-field generated. It is natural to define independence of random variables using the corresponding \( \sigma \)-fields, since the \( \sigma \)-field contains all information about the random variable.

Definition 3.8 Two random variables \( X \) and \( Y \) are independent if \( \sigma(X) \) and \( \sigma(Y) \) are independent.

Definition 3.9 A family of random variables \( \{ X_i \mid i \in I \} \) are independent if \( \{ \sigma(X_i) \mid i \in I \} \) are independent.

Example 3.0.21 Let \( A, B \) are independent events iff \( \sigma(A) \) and \( \sigma(B) \) are independent \( \sigma \)-fields iff the random variables \( I_A \) and \( I_B \) are independent.

Proof. Let \( A \) and \( B \) are independent. Consider

\[
P(AB^c) = P(A) - P(AB) = P(A)(1 - P(B)) = P(A)P(B^c)
\]

Hence \( A \) and \( B^c \) are independent. Changing the roles of \( A \) and \( B \) we have \( A^c \) and \( B \) are independent. Now \( A^c \) and \( B \) are independent implies that \( A^c \) and \( B^c \) are independent.

Since

\[
\sigma(A) = \{ \emptyset, A, A^c, \Omega \}, \quad \sigma(B) = \{ \emptyset, B, B^c, \Omega \},
\]

it follows that \( \sigma(A) \) and \( \sigma(B) \) are independent. Converse statement is obvious.

The second part follows from
\[ \sigma(A) = \sigma(I_A), \sigma(B) = \sigma(I_B). \]

**Example 3.0.22** The trivial \( \sigma \)-field \( \mathcal{F}_0 = \{\emptyset, \Omega\} \) is independent of any \( \sigma \)-field of subsets of \( \Omega \).

**Example 3.0.23** Define a probability space \( (\Omega, \mathcal{F}, P) \) as follows:

\[
\Omega = \{(a_1, a_2, \ldots, a_n) | a_i = 0, 1; i = 1, \ldots, n\}, \quad \mathcal{F} = \mathcal{P}(\Omega)
\]

and

\[
P\{(a_1, a_2, \ldots, a_n)\} = \frac{1}{2^n} \text{ for all } (a_1, \ldots, a_n) \in \Omega.
\]

For \( i = 1, 2, \ldots, n \), set

\[ A_i = \{(a_1, a_2, \ldots, a_n) | a_i = 1\}. \]

Then \( A_1, A_2, \ldots, A_n \) are independent.

This can be seen from the following.

\[
P(A_i) = \frac{2^{n-1}}{2^n} = \frac{1}{2} \quad \forall \ i = 1, 2, \ldots, n.
\]

Note that

\[
A_{i_1} \ldots A_{i_m} = \{(a_1, \ldots, a_m) | a_{i_1} = \cdots = a_{i_m} = 1\}.
\]

Hence

\[
P(A_{i_1} \ldots A_{i_m}) = \frac{2^{n-m}}{2^n} = \frac{1}{2^m} \quad \forall \ i_1, \ldots, i_m.
\]

Also

\[
P(A_{i_1}) \ldots P(A_{i_m}) = \left(\frac{1}{2}\right)^m = \frac{1}{2^m}.
\]

**Example 3.0.24** Let \( (\Omega, \mathcal{F}, P) \) is given by as

\[
\Omega = \{HH, HT, TH, TT\}, \quad \mathcal{F} = \mathcal{P}(\Omega)
\]

and
Define two random variables $X_1, X_2$ as

\[ X_1(HH) = X_1(HT) = 1, \quad X_1(TH) = X_1(TT) = 0, \]
\[ X_2(HH) = X_2(TH) = 1, \quad X_2(HT) = X_2(TT) = 0. \]

Then $X_1$ and $X_2$ are independent. Here note that

\[ \sigma(X_1) = \{ \emptyset, \{HH, HT\}, \{TH, TT\}, \Omega \} \]

and

\[ \sigma(X_2) = \{ \emptyset, \{HH, TH\}, \{HT, TT\}, \Omega \}. \]

We conclude this chapter with the celebrated Borel-Cantelli Lemma. To this end, we need the notion of limsup and liminf of events.

**Definition 3.10** (lim sup of sets) For $A_1, A_2, \ldots$, subsets of $\Omega$, define

\[ \limsup_n A_n := \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k. \]

Similarly

\[ \liminf_n A_n = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty}. \]

In the following theorem, we list some useful properties of limsup and liminf. This will make the objects \( \liminf \) and \( \limsup \) of sets much clearer.

**Theorem 3.0.11**

1. Let

\[ \{A_n \text{ i.o.}\} := \{\omega \in \Omega \mid \omega \in A_n \text{ for infinitely many } n\}. \]

Then

\[ \limsup_n A_n = \{A_n \text{ i.o.}\}. \]

2. Let

\[ \{A_n \text{ all but finitely many}\} := \{\omega \in \Omega \mid \omega \in A_n \text{ except for finitely many } n\}. \]

Then

\[ \liminf_n A_n = \{A_n \text{ all but finitely many}\}. \]
3. The following inclusion holds.
\[ \limsup_{n} A_n \supseteq \liminf_{n} A_n. \]

4. The following identity holds.
\[ [\limsup_{n} A_n]^c = \liminf_{n} (A_n)^c. \]

5. If \( A_1 \subseteq A_2 \subseteq \cdots \), then
\[ \limsup_{n} A_n = \bigcup_{n=1}^{\infty} A_n = \liminf_{n} A_n \]

6. If \( A_1 \supseteq A_2 \supseteq \cdots \), then
\[ \limsup_{n} A_n = \bigcap_{n=1}^{\infty} A_n = \liminf_{n} A_n. \]

**Proof.**

1. Consider
\[ \omega \in \limsup_{n} A_n \iff \omega \in \bigcup_{k=n}^{\infty} A_k \text{ for all } n \neq 1 \]
\[ \iff \text{There exists } n_1, n_2, \ldots \text{ such that } \omega \in A_{n_k}, \text{ for all } k \geq 1 \]
\[ \iff \omega \in A_n \text{ i.o.} \]

This proves (1).

2. Consider
\[ \omega \in \liminf_{n} A_n \iff \omega \in \bigcap_{k=n}^{\infty} A_k \text{ for some } n \neq 1 \]
\[ \iff \text{There exists } n_0 \geq 1 \text{ such that } \omega \in A_n, \text{ for all } n \geq n_0 \]
\[ \iff \omega \in A_n \text{ all but finitely many} \]

Thus (2).

3. Proof of (3) follows from (1) and (2).

4. Proof of (4) follows from De Morgan's laws.

5. Consider
\[ \limsup_{n} A_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k = \bigcap_{n=1}^{\infty} (\bigcup_{k=n}^{\infty} A_k) = \bigcup_{n=1}^{\infty} A_n \text{ (since } A_1 \subseteq A_2 \subseteq \cdots). \]

Similarly
\[ \liminf_{n} A_n = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k = \bigcup_{n=1}^{\infty} A_n. \]

This proves (5).
6. The proof of (6) follows from (4) and (5).

**Remark 3.0.4** The properties (1)-(6) are analogous to the corresponding properties of \( \limsup \) and \( \liminf \) of real numbers.

**Remark 3.0.5** Analogous to the notion of limit of sequence of numbers, one can say that \( \lim_{n \to \infty} A_n \) exists if

\[
\liminf_{n \to \infty} A_n = \limsup_{n \to \infty} A_n.
\]

Hence from Theorem 3.0.11 (5), if \( \{A_n\} \) is an increasing sequence of sets, then \( \lim_{n \to \infty} A_n \) exists and is \( \bigcup_{n=1}^{\infty} A_n \). Similarly using Theorem 3.0.11 (6), if \( \{A_n\} \) is a decreasing sequence of sets, then \( \lim_{n \to \infty} A_n \) exists and is \( \bigcap_{n=1}^{\infty} A_n \). Now student can see why property (6) in Theorem 1.0.1 is called the continuity property of probability.

**Theorem 3.0.12** (Borel - Cantelli Lemma) Let \( (\Omega, \mathcal{F}, P) \) be a probability space and \( A_1, A_2, \ldots \in \mathcal{F} \).

(i) If

\[
\sum_{n=1}^{\infty} P(A_n) < \infty
\]

then

\[
P(\limsup_{n} A_n) = 0.
\]

(ii) If

\[
\sum_{n=1}^{\infty} P(A_n) = \infty
\]

and

\[\{A_n \mid n = 1, 2, \ldots\}\]

are independent, then

\[
P(\limsup_{n} A_n) = 1.
\]

**Proof:** (i) Consider
\[ P(\limsup_{n} A_{n}) = P(\cap_{n=1}^{\infty} \cup_{k=n}^{\infty} A_{k}) \]
\[ \leq P(\cup_{k=n}^{\infty} A_{k}) \forall n = 1, 2, \ldots \]
\[ \leq \sum_{k=n}^{\infty} P(A_{k}) \text{for all } n = 1, 2, \ldots. \]

Note that the r.h.s $\to 0$ as $n \to \infty$, since
\[ \sum_{n=1}^{\infty} P(A_{n}) < \infty. \]

Therefore
\[ P(\limsup_{n} A_{n}) = 0. \]

(ii) Note that

\[ (\limsup_{n} A_{n})^{c} = \cup_{n=1}^{\infty} \cap_{k=n}^{\infty} A_{k}^{c}. \]

Therefore
\[ P(\limsup_{n} A_{n}) = 1 - P(\cup_{n=1}^{\infty} \cap_{k=n}^{\infty} A_{k}^{c}). \]

Now
\[ P(\cap_{k=n}^{\infty} A_{k}^{c}) \leq P(\cap_{k=n}^{n+m} A_{k}^{c}) \text{ for all } m = 1, 2, \ldots \]
\[ = \Pi_{k=n}^{n+m} P(A_{k}^{c}). \]

The last equality follows from
\[{A_1, A_2, \cdots \text{ are independent} \Rightarrow \{A_1^{c}, A_2^{c}, \cdots \text{ are independent.}}\]

Using $1 - x \leq e^{-x}$, we get
\[ P(\cap_{k=n}^{\infty} A_{k}^{c}) \leq \Pi_{k=n}^{n+m} (1 - P(A_{k})) \leq \Pi_{k=n}^{n+m} e^{-P(A_{k})}. \]

Since
\[ \sum_{n=1}^{\infty} P(A_{n}) = \infty \]

we get
\[ \lim_{n \to \infty} e^{-\sum_{k=n}^{n+m} P(A_{k})} = 0. \]

Therefore
\[ P(\cap_{k=n}^{\infty} A_{k}^{c}) = 0 \forall n = 1, 2, \ldots. \]
Thus
\[ P(\bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k^c) \leq \sum_{n=1}^{\infty} P(\bigcap_{k=n}^{\infty} A_k^c) = 0. \]

Therefore
\[ P(\limsup_{n} A_n) = 1. \]

This completes the proof.

**Example 3.0.25** Define a probability space \((\Omega, \mathcal{F}, P)\) as follows.

\[ \Omega = \{(a_1, a_2, \cdots, a_n, \cdots) | a_n = 0, 1; n = 1, 2, \cdots \}. \]

Set
\[ \mathcal{I} = \{A_{r_1 r_2 \cdots r_n} | n = 1, 2, \cdots; r_i = 0, i = 1, 2, \cdots n\} \]

where
\[ A_{r_1 r_2 \cdots r_n} = \{(a_1, a_2, \cdots, a_n, \cdots) | a_i = r_i, i = 1, 2, \cdots, n\}. \]

Let
\[ \mathcal{B}_0 = \text{Set of all finite disjoint union of members of } \mathcal{I}. \]

Then \(\mathcal{B}_0\) is a field, see Exercise 3.5. Define
\[ \mathcal{F} = \sigma(\mathcal{B}_0). \]

For \(A_{(r_1 \cdots r_n)} \in \mathcal{I},\) define
\[ P(A_{r_1 \cdots r_n}) = \frac{1}{2^n}. \]

For \(B \in \mathcal{B}_0,\) there exists \(I_i \in \mathcal{I}, i = 1, 2, \cdots, n\) which are pairwise disjoint such that
\[ B = \bigcup_{i=1}^{n} I_i. \]

Define \(P\) as follows.
\[ P(B) = P(\bigcup_{i=1}^{n} I_i) = \sum_{i=1}^{m} P(I_i). \]

Now extend \(P\) to \(\mathcal{F}\) using the extension theorem.

Let
\[ A = \{(a_1, a_2, \cdots) | a_n = 1 \text{ for infinitely many } n's\}. \]

One can calculate \(P(A)\) using Borel - Cantelli Lemma as follows. Define
\[ A_n = \{(a_1, a_2, \ldots) | a_n = 1\}, n = 1, 2, \ldots \]

Then

\[ \{A_1, A_2, \ldots\} \]

are independent and

\[ P(A_n) = \frac{1}{2} \text{ for all } n = 1, 2, \ldots. \]

Hence

\[ \sum_{n=1}^{\infty} P(A_n) = \infty \]

Also note that

\[ \limsup A_n = A. \]

Therefore using Borel - Cantelli Lemma

\[ P(A) = 1. \]