Chapter 3
Conditional Expectations

When we work on a particular branch of multi-period Binomial model, conditional probability is involved: conditional on a node, we consider the probability of going up or down. Given a continuous time price process $S_t$, if at time $u$ we want to find a product’s price at maturity $T$, we need to deal with the conditional distribution of $S_T$ given $S_u$. In this section we provide a rigorous introduction to conditional probabilities and conditional expectations.

3.1 Conditional Probability and Independence

**Definition 3.1. (Conditional Probability)** For any events $A, B \in \mathcal{F}$ such that $P(B) \neq 0$, the conditional probability of $A$ given $B$ is defined by

$$P(A|B) = \frac{P(A \cap B)}{P(B)}.$$  (3.1)

**Example 3.1. (Total Probability Formula).** For any event $A \in \mathcal{F}$ and any partition $B_1, B_2, \ldots$ of $\Omega$ such that $P(B_i) \neq 0$ and $B_i \in \mathcal{F}$ for any $i$,

$$P(A) = \sum_{i=1}^{\infty} P(A \cap B_i) = \sum_{i=1}^{\infty} P(A|B_i)P(B_i),$$

where the first and second equalities follow from the countably additivity of probability measure and (3.1) respectively.

**Definition 3.2. (Independence of events)** Two events $A, B \in \mathcal{F}$ are called independent if $P(A \cap B) = P(A)P(B)$. Any $n$ events $A_1, A_2, \ldots, A_n \in \mathcal{F}$ are independent if

$$P(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}) = P(A_{i_1})P(A_{i_2}) \cdots P(A_{i_k})$$

for any indices $1 \leq i_1 < i_2 < \cdots < i_k \leq n$. In general, an infinite family of events is independent if any finite number of them are independent.
**Definition 3.3. (Independence of random variables)** A set of $n$ random variables $X_1, \ldots, X_n$ are **independent** if for any Borel sets $B_1, \ldots, B_n \in \mathcal{B}$, the events 

$$\{X_1 \in B_1\}, \ldots, \{X_n \in B_n\}$$

are independent. In general, an **infinite** family of random variable is **independent** if any **finite** number of them are independent.

**Definition 3.4. (Independence of $\sigma$-fields)** A set of $n$ $\sigma$-fields $\mathcal{G}_1, \ldots, \mathcal{G}_n$ contained in $\mathcal{F}$ are **independent** if any **finite** number of them are independent. In general, an infinite family of $\sigma$-fields is **independent** if any **finite** number of them are independent.

**Remark 3.1.** The $\sigma$ fields $\mathcal{G}_i$s contained in $\mathcal{F}$, i.e., $\mathcal{G}_i \subset \mathcal{F}$ for all $i$. Otherwise, the measure $\mathbb{P}$ is not able to assign probabilities on the set $\{B_i \in \mathcal{G}_i\}$.

**Example 3.2.** Consider the dice example $(\Omega, \mathcal{F}, \mathbb{P})$ where $\Omega = \{1, 2, 3, 4, 5, 6\}$ and $\mathcal{F} = 2^\Omega$. Let $\mathbb{P}(\{\omega\}) = 1/6$ for $\omega = 1, \ldots, 6$. and 

$$A_1 = \{1, 2, 3\}, A_2 = \{2, 5\}, X_i = 1_{A_i}, \mathcal{F}_i = \sigma(\{A_i\}) \text{ for } i = 1, 2.$$ 

First consider independence of events: $\mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_1)\mathbb{P}(A_2) = \frac{1}{6}$, thus $A_1$ and $A_2$ are independent.

Next consider independence of r.v.s. Note that the $\sigma$-field generated by $X_1$ and $X_2$ are $\{\emptyset, A_1, \{4, 5, 6\}, \Omega\}$ and $\{\emptyset, A_2, \{1, 3, 4, 6\}, \Omega\}$ respectively. We can conclude that $X_1$ and $X_2$ are independent if each pair of these two groups of events are independent (4×4=16 combinations). Nevertheless, these 16 pairs of events can be shown to be independent by elementary calculations. Thus, $X_1$ and $X_2$ are independent.

Finally, to show that $\mathcal{F}_1$ and $\mathcal{F}_2$ are independent, we need to show that each combination of events from $\mathcal{F}_1$ and $\mathcal{F}_2$ are independent. Notice that $\mathcal{F}_i$ coincide with the $\sigma$-field generated by $X_i$ for $i = 1, 2$, respectively. Thus, the calculations in the preceding paragraph show that $\mathcal{F}_1$ and $\mathcal{F}_2$ are indeed independent.

From the above, it is seen that if two $\sigma$-fields $\mathcal{F}_a$ and $\mathcal{F}_b$ are independent, any $\mathcal{F}_a$ measurable r.v. is independent of any $\mathcal{F}_b$ measurable r.v..

### 3.2 Conditional Expectation

#### 3.2.1 Conditioning on an Event

**Definition 3.5. (Conditional Expectation given an Event)** For any random variable $X \in \mathcal{L}^1$ on $(\Omega, \mathcal{F}, \mathbb{P})$ and any event $B \in \mathcal{F}$ such that $\mathbb{P}(B) \neq 0$, the **conditional**
The number along an arrow represents the probability that the event indicated by the arrow would occur. Let $X$ be the value of the asset at the end of the second year. Recall the definition of path in Definition 1.6 where $\omega_t = (z_1, \ldots, z_t)$, $z_i = \pm 1$ for $i = 1, \ldots, t$, indicates the evolution (up=1, down=-1) of the market. We can take the probability space to be $(\Omega_X, \mathcal{F}_X, \mathbb{P}_X)$ where $\Omega_X = \{ \omega_2 = (a, b) | a = \pm 1, b = \pm 1 \}$, $\mathcal{F}_X = 2^{\Omega_X}$ and $\mathbb{P}_X$ is the discrete probability measure satisfying $\mathbb{P}_X(\{(1, 1)\}) = 1/6, \ldots, \mathbb{P}_X(\{(-1, -1)\}) = 1/4$.

Consider the event $B = \{ \omega_1 = (1) \}$. Clearly, $\mathbb{P}(B) = 2/3$. Also, note that

$$X((1, 1)) = 12.1, \quad X((1, -1)) = 10.45.$$ 

and

$$\mathbb{P}(\{(1, 1)\}) = \frac{1}{6}, \quad \mathbb{P}(\{(1, -1)\}) = \frac{1}{2}.$$ 

Therefore

$$E(X|B) = \frac{1}{\mathbb{P}(B)} \int_B X d\mathbb{P} = \frac{1}{\mathbb{P}(B)} \left[ (X((1, 1))\mathbb{P}(\{(1, 1)\}) + X((1, -1))\mathbb{P}(\{(1, -1)\})) \right]$$

$$= \frac{1}{2/3} \left( 12.1 \times \frac{1}{6} + 10.45 \times \frac{1}{2} \right) = 10.8625.$$ 

$\square$
3.2.2 Conditioning on a discrete random variable

Given an arbitrary random variable $X \in \mathcal{L}^1$ (so expectation exists) and a discrete random variable $Z : \Omega \to \{ z_1, \ldots, z_m \}$, the conditional expectation of a r.v. $X$ given $Z$, i.e. $E(X|Z)$, must depend solely on the random variable $Z$. Thus $E(X|Z)$ is itself a random variable. Since $Z$ takes only $m$ possible values, so does $E(X|Z)$. Therefore, we can characterize $E(X|Z)$ by $m$ conditional expectations (on an event $\{ Z = z_i \}$) $E(X|\{ Z = z_i \})$, $i = 1, \ldots, m$, using the definition in the last section.

**Definition 3.6. (Conditional Expectation given a discrete r.v.)** Let $X \in \mathcal{L}^1$ and $Z$ be a discrete r.v. taking values on $\{ z_i \}_{i=1}^{m}$. The conditional expectation of $X$ given $Z$ is defined to be a discrete random variable $E(X|Z) : \Omega \to \mathbb{R}$ such that

$$E(X|Z)(\omega) = E(X|\{ Z = z_i \}) \quad \text{on} \quad \{ \omega : Z(\omega) = z_i \}, \quad (3.2)$$

for any $i = 1, 2, \ldots$ □

**Example 3.4.** Three coins, 10, 20 and 50 cents, are tossed. The values of those coins with heads up are added for a total amount $X$. Let $Z$ be the total amount of the 10 and 20 cents, what is $E(X|Z)$?

First, the probability space can be taken as $\Omega = \{ HHH, HHT, \ldots, TTT \}$, where $H$ and $T$ stands for Head and Tail respectively. The associated $\sigma$-field can be taken as $\mathcal{F} = 2^\Omega$. The probability measure $\mathbb{P}$ may be defined on $\mathcal{F}$ with $\mathbb{P}(\{ HHH \}) = \cdots = \mathbb{P}(\{ TTT \}) = \frac{1}{8}$. The discrete r.v. $Z$ on $(\Omega, \mathcal{F}, \mathbb{P})$ is given by $Z : \Omega \to \{ 0, 10, 20, 30 \}$. Thus finding $E(X|Z)$ is equivalent to finding $E(X|\{ Z = 0 \}), \ldots, E(X|\{ Z = 30 \})$. For example,

$$E(X|\{ Z = 0 \}) = \frac{1}{\mathbb{P}(\{ Z = 0 \})} \int_{Z=0} X \mathbb{P} = \frac{1}{1/4} \left( (0+50) \frac{1}{8} + (0+0) \frac{1}{8} \right) = 25.$$  

Using similar argument for the other cases, we have

$$E(X|Z)(\omega) = \begin{cases} 25 & \text{if } Z(\omega) = 0, \\ 35 & \text{if } Z(\omega) = 10, \\ 45 & \text{if } Z(\omega) = 20, \\ 55 & \text{if } Z(\omega) = 30. \end{cases}$$

Note that the probability measure on $Z$ is irrelevant. In other words, the above results still hold even if the 10 and 20 cents coins are biased. However, the results will be different if the 50 cents coin is assumed biased. □

**Theorem 3.1.** If $X \in \mathcal{L}^1$ and $Z$ is a discrete random variable, then

i) $E(X|Z)$ is $\sigma(Z)$-measurable

ii) For any $A \in \sigma(Z)$,

$$\int_A E(X|Z) d\mathbb{P} = \int_A X d\mathbb{P}. \quad (3.3)$$
Proof. i) To show that \(E(X|Z)\) is \(\sigma(Z)\) measurable, we have to show that for any \(B \in \mathcal{B}, \{\omega : E(X|Z)(\omega) \in B\} \in \sigma(Z)\). Since \(Z\) is discrete, we can assume that \(Z\) takes values \(\{z_i\}_{i=1}^{\infty}\). From (3.2), \(E(X|Z)\) also takes discrete values \(\{y_i\}_{i=1}^{\infty}\), where \(y_i = E(X|\{Z = z_i\})\). Note that \(\{\omega : E(X|Z)(\omega) = y_i\} = \{Z = z_i\}\). It follows that for any \(B \in \mathcal{B}, \{\omega : E(X|Z)(\omega) \in B\} \) is a disjoint union of \(\{Z = z_i\}\), which belongs to \(\sigma(Z)\). This completes the proof of i).

ii) Since \(E(X|Z)(\omega)\) is constant on \(\{\omega : Z(\omega) = z_i\}\), we have

\[
\int_{\{Z = z_i\}} E(X|Z)dP = \int_{\{Z = z_i\}} E(X|\{Z = z_i\})dP
\]

\[
= E(X|\{Z = z_i\}) \int_{\{Z = z_i\}} dP
\]

\[
= \frac{\int_{\{Z = z_i\}} XdP}{\int_{\{Z = z_i\}} dP} \int_{\{Z = z_i\}} dP
\]

\[
= \int_{\{Z = z_i\}} XdP.
\]

In general, any \(A \in \sigma(Z)\) is a countable union of sets of the form \(\{Z = z_i\}\), which are pairwise disjoint. Thus (3.3) follows from the countably additivity of Lebesgue integral.

**Example 3.5.** Continuing Example 3.4, the induced sample space for the r.v. \(X\) and \(Z\) are \(\Omega_X = \{0, 10, 20, 30, 50, 60, 70, 80\}\) and \(\Omega_Z = \{0, 10, 20, 30\}\). Let \(A_1 = \{Z = 20\}\) and \(A_2 = \{Z = 10\} \cup \{Z = 30\}\). Note that \(A_1, A_2 \in \mathcal{F}_Z \triangleq 2^{\Omega_Z}\). It is easy to verify that (3.3) holds for \(A = A_1\) and \(A_2\). In particular,

\[
\int_{A_1} XdP = 20 \frac{1}{8} + 70 \frac{1}{8} = 11.25,
\]

\[
\int_{A_1} E(X|Z)dP = 45 \frac{1}{4} = 11.25,
\]

\[
\int_{A_2} XdP = (10 + 60 + 30 + 80) \frac{1}{8} = 22.5,
\]

\[
\int_{A_2} E(X|Z)dP = (35 + 55) \frac{1}{4} = 22.5.
\]

\[\square\]

3.2.3 Conditioning on an arbitrary random variable

If \(Z\) is a discrete r.v., we see from (3.2) that \(E(X|Z)\) is also a discrete random variable. When \(Z\) is an arbitrary r.v. (has discrete and continuous component), it is not easy to write the explicit formula for \(E(X|Z)\) in general. (The elementary way of
writing \( E(X|Z) = \int x f_{X|Z}(x,z)/f_Z(z) \, dx \) cannot be used if the p.d.f.s \( f_{X|Z} \) or \( f_Z(z) \) does not exist.) However, Theorem 3.1 suggests a way for us to define conditional expectation given an arbitrary r.v..

**Definition 3.7. (Conditional Expectation given an arbitrary r.v.)** Let \( X \in \mathcal{L}^1 \) and \( Z \) be an arbitrary r.v.. Then the **conditional expectation** of \( X \) given \( Z \) is defined to be a random variable \( E(X|Z) \) such that

i) \( E(X|Z) \) is a \( \sigma(Z) \)-measurable r.v.,

ii) For any \( A \in \sigma(Z) \),

\[ \int_A E(X|Z) \, d\mathbb{P} = \int_A X \, d\mathbb{P}. \tag{3.4} \]

Since conditional expectation is defined implicitly by (3.4) rather than an explicit formula, we need to justify such a definition by showing that the r.v. \( E(X|Z) \) is characterized uniquely. Since \( E(X|Z) \) is a random variable, the uniqueness depends on the probability measure \( \mathbb{P} \) through the notion of **almost sure equivalence**.

**Definition 3.8. (Almost Sure equivalence)**

1. Two events \( A, B \in \mathcal{F} \) are equal almost surely (a.s.) if \( \mathbb{P}((A \setminus B) \cup (B \setminus A)) = 0 \), i.e., the un-common region of \( A \) and \( B \) has probability 0.

2. Two random variables \( X \) and \( Y \) on \( (\Omega, \mathcal{F}, \mathbb{P}) \) are equal almost surely (a.s.) if \( \mathbb{P}(X = Y) = \mathbb{P}(\{ \omega \in \Omega : X(\omega) = Y(\omega) \}) = 1 \).

**Example 3.6.** Note that two events \( A \) and \( B \) are equal a.s. does not imply that the two events are the same. It only means that the event that \( A \setminus B \) and \( B \setminus A \) are of \( \mathbb{P} \)-measure 0. For example, consider \( A = (0,0.5) \) and \( B = (0,0.5) \) on \( (\{0,1\}, \mathcal{B}_{\{0,1\}}, \lambda_{\{0,1\}}) \). It is clear that \( A = B \) a.s. but \( A \neq B \).

For the case of random variables, let \( \Omega = \{ \omega_1, \omega_2 \} \), \( \mathcal{F} = 2^\Omega \), \( \mathbb{P} \) is defined on \( \mathcal{F} \) and satisfies \( \mathbb{P}(\{ \omega_1 \}) = 1 \). Suppose that \( X_1, X_2 : \Omega \to \mathbb{R} \) are r.v.s with \( X_1 = \mathbb{I}_{\omega_1} + 2 \times \mathbb{I}_{\omega_2} \) and \( X_2 = \mathbb{I}_{\omega_1} + 3 \times \mathbb{I}_{\omega_2} \). Then \( X_1 = X_2 \) a.s. but \( X_1 \neq X_2 \). \( \square \)

**Theorem 3.2. (Uniqueness of Conditional Expectation)** \( E(X|Z) \) is unique in the sense that if \( X = X' \) a.s., then \( E(X|Z) = E(X'|Z) \) a.s.

The proof of Theorem 3.2 relies on the following lemma.

**Lemma 3.3** Let \( (\Omega, \mathcal{F}, \mathbb{P}) \) be a probability space and let \( \sigma(Z) \in \mathcal{F} \). If \( Y \) is a \( \sigma(Z) \) measurable r.v. and

\[ \int_A Y \, d\mathbb{P} = 0, \]

for any \( A \in \sigma(Z) \), then \( Y = 0 \) a.s. That is, \( \mathbb{P}(Y = 0) = 1 \).

**Proof.** Observe that \( P\{ Y \geq \frac{1}{n} \} = 0 \) for any positive integer \( n \) because

\[ 0 \leq \frac{1}{n} \mathbb{P}\{ Y \geq \frac{1}{n} \} = \int_{\{Y \geq \frac{1}{n} \}} \frac{1}{n} \, d\mathbb{P} \leq \int_{\{Y \geq \frac{1}{n} \}} Y \, d\mathbb{P} = 0. \]
The last equality follows from the fact that \( \{ Y \geq \frac{1}{n} \} = \{ Y \in [\frac{1}{n}, \infty) \} \in \sigma(Z) \). Similarly, \( P\{ Y \leq -\frac{1}{n} \} = 0 \) for any \( n > 0 \). Set \( A_n = \{ -\frac{1}{n} < Y < \frac{1}{n} \} \), then we have

\[
P(A_n) = 1
\]

for any positive integer \( n \). Since \( \{ A_n \} \) forms a contracting sequence of events and \( \{ Y = 0 \} = \bigcap_{n=1}^{\infty} A_n \), it follows that

\[
P\{ Y = 0 \} = P \left( \bigcap_{n=1}^{\infty} A_n \right) = \lim_{m \to \infty} P \left( \bigcap_{n=1}^{m} A_n \right) = \lim_{m \to \infty} P(A_m) = 1, \tag{3.5}
\]

as required. See Exercise 3.29.

**Proof. (Theorem 3.2.)** Suppose that \( X = X' \) a.s., i.e., \( P(X = X') = 1 \). We want to show that \( E(X|Z) = E(X'|Z) \) a.s. Let \( Y = E(X|Z) - E(X'|Z) \), note that by the definition of conditional distribution, \( E(X|Z) \) and \( E(X'|Z) \) are \( \sigma(Z) \) measurable, so does \( Y \). For any \( A \in \sigma(Z) \), we have

\[
\int_A Y \, dP = \int_A (E(X|Z) - E(X'|Z)) \, dP
\]

\[
= \int_A (X - X') \, dP \quad \text{(by (3.4))}
\]

\[
= 0 \quad \text{(Since \( X = X' \) a.s.)}
\]

From Lemma 3.3, we have \( Y = 0 \) a.s., i.e., \( E(X|Z) = E(X'|Z) \) a.s., completing the proof of Theorem 3.2.

**Example 3.7.** Consider the probability space \((\mathbb{R}, B_{[0,1], \mathbb{P}_{[0,1]}})\). Let

\[
X(\omega) = 2\omega^2, \quad Z(\omega) = \begin{cases} 2 & \text{if } \omega \in [0, \frac{1}{2}) \\ \omega & \text{if } \omega \in [\frac{1}{2}, 1] \end{cases}.
\]

The following heuristic can be used to guess the form of \( E(X|Z) \): First note that there is a discrete mass at \( Z = 2 \), we can follow the arguments used in conditioning on discrete r.v. to obtain that, on \( \omega \in [0, \frac{1}{2}) \),

\[
E(X|Z)(\omega) = E(X|\{Z = 2\}) = \frac{1}{\mathbb{P}(\omega \in [0, \frac{1}{2}))} \int_{[0, \frac{1}{2})} X(\omega) \, d\mathbb{P}
\]

\[
= \frac{1}{1/2} \int_0^{\frac{1}{2}} 2\omega^2 \, d\omega = \frac{1}{6}. \tag{3.6}
\]
For the continuous part $Z(\omega) = \omega$ on $\omega \in [1/2, 1]$, note that \{\(Z = z\)\} $\in \sigma(Z)$ and \{\(Z = z\)\} = \{\(\omega : Z(\omega) = z\)\} = \{\(\omega = z\)\} for $z \in [1/2, 1]$, we heuristically solve for $E(X|Z)$ from the definition by

$$
\int_{\{Z = z\}} E(X|Z) d\mathbb{P} = \int_{\{Z = z\}} X d\mathbb{P} \\
\Rightarrow E(X|Z = z) \int_{\{Z = z\}} d\mathbb{P} = \int_{\{Z = z\}} X \mathbb{P}(d\omega) \\
\Rightarrow E(X|Z = z) = \frac{\int_X X d\mathbb{P}}{\mathbb{P}(\{Z = z\})} = X(z) = 2z^2.
$$

(3.7)

Express $E(X|Z)$ as a function of $\omega$ (sample space) instead of a function of $Z$ (induced sample space), we have

$$
E(X|Z)(\omega) = \begin{cases} 
\frac{1}{6} & \text{if } \omega \in [0, \frac{1}{2}) \\
2\omega^2 & \text{if } \omega \in [\frac{1}{2}, 1]. 
\end{cases}
$$

To verify that $E(X|Z)$ is really a conditional expectation, we need to verify Conditions i) and ii) in Definition 3.7. First we show i): $E(X|Z)$ is $\sigma(Z)$ measurable. Recall that $\sigma(Z) = \{Z^{-1}(B), B \in \mathcal{B}\}$. For any Borel set $B \in \mathcal{B}$, from the definition of $Z$, we have

$$
Z^{-1}(B) = \begin{cases} 
B \cap [\frac{1}{2}, 1] & \text{if } 2 \notin B, \\
(B \cap [\frac{1}{2}, 1]) \cup [0, \frac{1}{2}) & \text{if } 2 \in B. 
\end{cases}
$$

(3.8)

On the other hand, it can be checked that the inverse mapping of \{\(\omega : E(X|Z)(\omega) \in B\)\}, where $B \in \mathcal{B}$, also takes the same form as in (3.8). Hence, \{\(\omega : E(X|Z)(\omega) \in B\)\} $\subseteq \{Z^{-1}(B), B \in \mathcal{B}\} = \sigma(Z)$, i.e., $E(X|Z)$ is $\sigma(Z)$ measurable. Finally, ii) can be directly justified by calculations similar to (3.6) and (3.7). Thus, $E(X|Z)$ satisfies Definition 3.7, i.e., is a valid conditional expectation.

Example 3.8. (Elementary definition of Conditional Expectation) Recall that Definition 3.7 of conditional expectation $E(X|Z)$ does not require the existence of joint p.d.f. of $X$ and $Z$. In this example we show that $E(X|Z)$ reduces to the elementary definition of conditional expectation when the joint p.d.f. of $X$ and $Z$ exists.

Suppose that the joint p.d.f. of r.v.s $X$ and $Z$ exists and is given by $f_{X,Z}(x,z)$. We can define the conditional p.d.f. by

$$
f_{X|Z}(x|z) = \frac{f_{X,Z}(x,z)}{f_Z(z)} \text{ if } f_Z(z) \neq 0, 
$$

and 0 otherwise, where $f_Z(z) = \int f_{X,Z}(x,z)dx$. Note that if the joint p.d.f. exists, then $d\mathbb{P} = f_{X,Z}(x,z)dxdz$. Following the same heuristic as in Example 3.7, we solve for $E(X|Z)$ from the definition by

$$
\int_{\{Z = z\}} E(X|Z) d\mathbb{P} = \int_{\{Z = z\}} X d\mathbb{P} \\
\Rightarrow E(X|Z = z) \int_{\{Z = z\}} d\mathbb{P} = \int_{\{Z = z\}} X f_{X,Z}(x,z)dx \\
\Rightarrow E(X|Z = z) = \frac{\int_X X f_{X,Z}(x,z)dx}{\int_X f_{Z}(z)} = \int_X X f_{X|Z}(x|z)dx = g(z),
$$

(3.9)

say. Then $E(X|Z) \triangleq g(Z)$ is the conditional expectation of $X$ given $Z$. 
Next we justify Conditions i) and ii) in Definition 3.7. First, i) is trivial since \(g(z)\) is a function involving only \(z\). Lastly we verify ii): for any \(B \in \mathcal{B}\) and \(A \triangleq \{ \omega : Z(\omega) \in B \} \in \sigma(Z)\), we have

\[
\int_A X \, d\mathbb{P} = \int \int I_A(z) x f_{XZ}(x,z) \, dx \, dz
\]

and

\[
\int_A g(Z) \, d\mathbb{P} = \int \int I_A(z) g(z) f_Z(z) \, dz = \int \int I_A(z) \left( \int x f_{X|Z}(x|z) \, dx \right) f_Z(z) \, dz
\]

\[
= \int \int I_A(z) x f_{XZ}(x,z) \, dx \, dz = \int_A X \, d\mathbb{P}.
\]

Thus conditions in Definition 3.7 are fulfilled.

\[\square\]

### 3.3 Conditioning on a \(\sigma\)-field

Note that Definition 3.7 only looks at sets of \(\sigma(Z)\) rather than on the actual values of \(Z\), we can generalize the definition of conditional expectation to a given \(\sigma\)-field.

**Definition 3.9.** Let \(X \in L^1\) on \((\Omega, \mathcal{F}, \mathbb{P})\), and let \(\mathcal{G} \subset \mathcal{F}\) be a \(\sigma\)-field. Then the **conditional expectation** of \(X\) given \(\mathcal{G}\) is defined to be a random variable \(E(X|\mathcal{G})\) such that

i) \(E(X|\mathcal{G})\) is a \(\mathcal{G}\)-measurable r.v.

ii) For any \(A \in \mathcal{G}\),

\[
\int_A E(X|\mathcal{G}) \, d\mathbb{P} = \int_A X \, d\mathbb{P}.
\]  

(3.10)

**Remark 3.2.** Combining Definition 3.7 and 3.9, it can be seen that

\[
E(X|\sigma(Z)) = E(X|Z).
\]

Note that different r.v.s can generate the same \(\sigma\)-field. For example, on the sample space \(\Omega = \{H,T\}\), let \(X_1 = 1_{\{H\}}\) and \(X_2 = 1_{\{T\}}\) generate the same \(\sigma\)-field \(\mathcal{F} = \{\emptyset, \{H\}, \{T\}, \Omega\}\). Thus, conditional expectation given a \(\sigma\)-field is more general than conditional expectation given a r.v..  

**Example 3.9.** If \(\mathcal{G} = \{\emptyset, \Omega\}\), then \(E(X|\mathcal{G}) = E(X)\) a.s. To see this, note that if \(\mathcal{G} = \{\emptyset, \Omega\}\), then any constant r.v., including \(E(X)\), is \(\mathcal{G}\)-measurable. Note that

\[
\int_{\Omega} X \, d\mathbb{P} = E(X) = \int_{\Omega} E(X) \, d\mathbb{P},
\]

and
\[
\int_{\emptyset} X \, dP = 0 = \int_{\emptyset} E(X) \, dP.
\]

Therefore, \(E(X)\) satisfies i) and ii) of Definition 3.7, thus \(E(X|\mathcal{G}) = E(X)\), a.s. \(\square\)

**Example 3.10.** For \(B \in \mathcal{G}\), we have \(E(E(X|\mathcal{G})|B) = E(X|B)\). To see this, note that from the definition of conditional expectation, for \(B \in \mathcal{G}\) we have

\[
\int_B E(X|\mathcal{G}) \, dP = \int_B X \, dP.
\]

Thus Definition 3.5 implies that

\[
E(E(X|\mathcal{G})|B) = \frac{1}{P(B)} \int_B E(X|\mathcal{G}) \, dP = \frac{1}{P(B)} \int_B X \, dP = E(X|B).
\]

**Example 3.11.** Consider the fair dice example where \(\Omega = \{1, \ldots, 6\}\) and \(P(\{i\}) = \frac{1}{6}\). Let \(\mathcal{F}_1 = \{\emptyset, \Omega\}\), \(\mathcal{F}_2 = \{\emptyset, \{1, 2, 3\}, \{4, 5, 6\}, \Omega\}\), \(\mathcal{F}_3 = 2^\mathcal{F}\) and \(X : \Omega \to \mathbb{R}\) satisfying \(X(i) = (i - 4)^+\). Note that

\[
\sigma(X) = \{\emptyset, \{5\}, \{6\}, \{5, 6\}, \{1, 2, 3, 4\}, \{1, 2, 3, 4, 5\}, \{1, 2, 3, 4, 6\}, \Omega\},
\]

thus \(X \in \mathcal{F}_3\) but \(X \notin \mathcal{F}_1\) and \(\mathcal{F}_2\). From Example 3.9, \(E(X|\mathcal{F}_1) = E(X) = (5 - 4)\frac{1}{6} + (6 - 4)\frac{1}{6} = 0.5\). Next, \(E(X|\mathcal{F}_2)\) is a \(\mathcal{F}_2\) measurable function, thus it takes constant values on each of \(\{1, 2, 3\}\) and \(\{4, 5, 6\}\). Simple calculations yield

\[
E(X|\mathcal{F}_2)(\omega) = \begin{cases} 
\frac{f_{\{1,2,3\}} X \, dP}{P(\{1,2,3\})} = 0 & \text{if } \omega \in \{1,2,3\}, \\
\frac{f_{\{4,5,6\}} X \, dP}{P(\{4,5,6\})} = 1 & \text{if } \omega \in \{4,5,6\}.
\end{cases}
\]

Lastly, since \(X \in \mathcal{F}_3\), we have \(E(X|\mathcal{F}_3) = X\).

### 3.4 Radon-Nikodym derivative

In Definition 3.7 and 3.9, it is implicitly assumed that the random variable \(E(X|Z)\) or \(E(X|\mathcal{G})\) exists. The existence of conditional expectation is non-trivial, but is guaranteed by the **Radon-Nikodym Theorem**, which is an important tool in probability theory that makes connections between measures.

**Definition 3.10.** **(Absolute Continuous measures)** Suppose that \(P\) and \(Q\) are probability measures on \((\Omega, \mathcal{F})\) and that \(P(F) = 0\) implies \(Q(F) = 0\) when \(F \in \mathcal{F}\). Then \(Q\) is said to be **absolutely continuous** with respect to \(P\) and is denoted by \(Q \ll P\).

**Example 3.12.** Consider the measurable space \((\Omega, \mathcal{F})\) where \(\Omega = \{1, 2, 3, 4, 5, 6\}\) and \(\mathcal{F} = 2^\mathcal{F}\). For \(\omega \in \Omega\), let \(P_1(\{\omega\}) = 1/6\), \(P_2(\{\omega\}) = |\omega - 3.5|/9\), \(P_3(\{\omega\}) = 1/4 1_{(\omega \geq 4)}\) and \(P_4(\{\omega\}) = 1/4 1_{(\omega \leq 3)}\). Then, the only absolute continuous pairs are \(P_1 \ll P_2\), \(P_2 \ll P_1\), \(P_3 \ll P_1\) and \(P_4 \ll P_1\) for \(i = 1, 2\).
For the measurable space \((\mathbb{R}, \mathcal{B})\), let \(\mathbb{P}_N\) and \(\mathbb{P}_E\) satisfies \(\mathbb{P}_N((a, b)) = F_N(b) - F_N(a)\) and \(\mathbb{P}_E((a, b)) = F_E(b) - F_E(a)\), where \(F_N\) and \(F_E\) are the c.d.f.s of the standard normal and standard exponential distribution family, respectively. Then we have \(\mathbb{P}_E \ll \mathbb{P}_N\), but \(\mathbb{P}_N \ll \mathbb{P}_E\) is not true.

\[\square\]

**Theorem 3.4. (Radon-Nikodym Theorem)** Let \(\mathbb{P}\) and \(\mathbb{Q}\) be two probability measures on the measurable space \((\Omega, \mathcal{F})\). If \(\mathbb{Q} \ll \mathbb{P}\), then there exists a \(\mathcal{F}\) measurable r.v. \(Y \in \mathcal{L}^1\) such that

\[
\mathbb{Q}(A) = \int_A Y \, d\mathbb{P},
\]

for any \(A \in \mathcal{F}\). The r.v. \(Y\) is called the **Radon-Nikodym derivative** of \(\mathbb{Q}\) w.r.t. \(\mathbb{P}\) and is denoted by \(Y = \frac{d\mathbb{Q}}{d\mathbb{P}}\).

**Proof.** The proof consists of four steps:

i) Assume \(\mathbb{Q}(A) \leq \mathbb{P}(A)\) for all \(A \in \mathcal{F}\). Construct a \(Y_n\) on a finite \(\sigma\)-field \(\mathcal{F}_n \subset \mathcal{F}\).

ii) Take limit to obtain the required r.v. \(Y = \lim_n Y_n\).

iii) Show that the limit \(Y\) satisfies (3.11) any \(A \in \mathcal{F}\).

iv) Relax the assumption \(\mathbb{Q}(A) \leq \mathbb{P}(A)\).

i) [Construction on finite \(\sigma\)-field] Assume that \(\mathbb{Q}(A) \leq \mathbb{P}(A)\) for all \(A \in \mathcal{F}\). First we define the required function \(Y\) on a finite sub-\(\sigma\)-field \(\mathcal{F}_n \subset \mathcal{F}\) which is generated by a finite partition of \(\Omega\). That is, for some integer \(r_n\), \(\Omega = \bigcup_{j=1}^{r_n} F_j\) where \(F_i \cap F_j = \emptyset\) if \(i \neq j\) and \(\mathcal{F}_n = \sigma(F_1, \ldots, F_{r_n})\). Define \(Y_n : \Omega \to \mathbb{R}\) by

\[
Y_n(\omega) = \begin{cases} 
\frac{\mathbb{Q}(F_j)}{\mathbb{P}(F_j)} & \text{if } \omega \in F_j \text{ and } \mathbb{P}(F_j) > 0, \\
0 & \text{if } \mathbb{P}(F_j) = 0.
\end{cases}
\]

(3.12)

Note that \(Y_n \in \mathcal{L}^1\) since

\[
\mathbb{E}(Y_n) = \int_\Omega Y_n \, d\mathbb{P} = \sum_{j=1}^{r_n} \int_{F_j} Y_n \, d\mathbb{P} = \sum_{j=1}^{r_n} \mathbb{Q}(F_j) = \mathbb{Q}(\Omega) = 1 < \infty.
\]

Moreover, \(Y_n(\omega)\) is a constant \((\mathbb{Q}(F_j)/\mathbb{P}(F_j))\) on \(\omega \in F_j\), so \(Y_n\) is \(\mathcal{F}_n\) measurable. Also, simple algebra shows that \(\mathbb{Q}(F_j) = \int_{F_j} Y_n \, d\mathbb{P}\). Since every \(F \in \mathcal{F}_n\) is a union of \(F_j\)'s, (3.11) holds for all \(A \in \mathcal{F}_n\).

Next consider a refinement \(\mathcal{F}_{n+1} = \sigma(\bar{F}_1, \ldots, \bar{F}_{r_{n+1}})\) of \(\mathcal{F}_n\) such that \(\{\bar{F}_j\}_{j=1}^{r_{n+1}}\) are disjoint and each \(F_i \in \mathcal{F}_n\) is a union of some \(\bar{F}_j\)’s. Define \(Y_{n+1} : \Omega \to \mathbb{R}\) by

\[
Y_{n+1}(\omega) = \begin{cases} 
\frac{\mathbb{Q}(\bar{F}_j)}{\mathbb{P}(\bar{F}_j)} & \text{if } \omega \in \bar{F}_j \text{ and } \mathbb{P}(\bar{F}_j) > 0, \\
0 & \text{if } \mathbb{P}(\bar{F}_j) = 0.
\end{cases}
\]

Similarly, it can be shown that \(Y_{n+1} \in \mathcal{L}^1\), \(Y_{n+1}\) is \(\mathcal{F}_{n+1}\) measurable, and \(\mathbb{Q}(A) = \int_A Y_{n+1} \, d\mathbb{P}\) for all \(A \in \mathcal{F}_{n+1}\). Moreover,
where the second equality follows from

\[
\int \Omega Y_{n}^{2} d\mathbb{P} = \int \Omega (Y_{n+1} - Y_n)^2 d\mathbb{P} + \int \Omega Y_n^2 d\mathbb{P} + 2 \int \Omega Y_n(Y_{n+1} - Y_n) d\mathbb{P}
\]

\[
= \int \Omega (Y_{n+1} - Y_n)^2 d\mathbb{P} + \int \Omega Y_n^2 d\mathbb{P}
\]

\[
\geq \int \Omega Y_n^2 d\mathbb{P},
\]

(3.13)

ii) [Passing to the limit] From (3.13), it can be observed that if the partition of \( \Omega \) gets finer and finer (e.g., a sequence of \( \sigma \)-field \( \mathcal{F}_1 \subset \cdots \subset \mathcal{F}_k \subset \cdots \)), then the corresponding Radon-Nikodym derivative \( Y_k \) is a non-decreasing sequence in the sense that \( \int \Omega Y_{k+1}^2 d\mathbb{P} \geq \int \Omega Y_k^2 d\mathbb{P} \). On the other hand, \( \int \Omega Y_k^2 d\mathbb{P} \) is bounded because \( \mathbb{P}(\Omega) = 1 \) and \( Y_k \leq 1 \) from (3.12) and the assumption \( \mathbb{Q}(A) \leq \mathbb{P}(A) \) for all \( A \in \mathcal{F} \).

Thus, the limit \( l := \lim_{k \to \infty} \int \Omega Y_k^2 d\mathbb{P} \) exists. Some limit arguments show that the limit \( Y = \lim_{k \to \infty} Y_k \) also exists (see Exercise 3.30).

iii) [Verify (3.11)] Since \( Y_n \in \mathcal{F}_n \subset \mathcal{F} \) and \( Y \) is the limit of \( Y_n \), \( Y \) is \( \mathcal{F} \) measurable.

For any \( A \in \mathcal{F} \), let \( \mathcal{A}_n \) be the common refinement of \( \mathcal{F}_n \) and \( \{A,A^c\} \). It can be verified directly that \( \mathbb{Q}(A) = \int_A Y_{\mathcal{A}_n} d\mathbb{P} \) where \( Y_{\mathcal{A}_n} \) is defined similarly as \( Y_n \) but is on \( \mathcal{A}_n \) instead of \( \mathcal{F}_n \). Similar limit arguments as in ii) show that \( \int_A Y_{\mathcal{A}_n} d\mathbb{P} \to \int_A Y d\mathbb{P} \) for some subsequence \( \{n_k\}_{k=1,2,...} \), yielding (3.11) (see Exercise 3.30).

iv) [Relax the assumption \( \mathbb{Q}(A) \leq \mathbb{P}(A) \)] In general, let \( S = \mathbb{P} + Q \), so \( \mathbb{P}(A) \leq S(A) \) and \( \mathbb{Q}(A) \leq S(A) \) for all \( A \in \mathcal{F} \). The previous results implies that there exist \( Y_Q \) and \( Y_P \) such that \( \mathbb{Q}(A) = \int_A Y_Q dS \) and \( \mathbb{P}(A) = \int_A Y_P dS \) for all \( A \in \mathcal{F} \). Now, let \( Y = \frac{Y_Q}{Y_P} 1_{\{Y_P > 0\}} \). Then,

\[
\mathbb{Q}(A) = \int_A Y_Q dS
\]

\[
= \int_{A \cap \{Y_P > 0\}} \frac{Y_Q}{Y_P} Y_P dS + \int_{A \cap \{Y_P = 0\}} Y_Q dS
\]

\[
= \int_A YY_P dS + 0
\]

\[
= \int_A Y d\mathbb{P},
\]
where the third equality follows from that on $B := \{ Y_0 = 0 \}$, $\mathbb{P}(B) = \int_B Y_0 d\mathcal{S} = 0$, which implies $Q(B) = 0$ as $Q \ll \mathbb{P}$, which in turn implies that $\mathcal{S}(B) = 0$. The last equality follows from Exercise 3.30. Thus the proof is completed.

Using the Radon-Nikodym Theorem, the existence of conditional expectation can be justified:

**Theorem 3.5. (Existence of Conditional Expectation)** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $\mathcal{G}$ be a $\sigma$-field contained in $\mathcal{F}$. Then for any r.v. $X \in \mathcal{L}^1$ there exists a $\mathcal{G}$-measurable r.v. $X_\mathcal{G}$ such that for every $A \in \mathcal{G}$,

$$
\int_A X d\mathbb{P} = \int_A X_\mathcal{G} d\mathbb{P}.
$$

Here $X_\mathcal{G}$ can be regarded as the conditional expectation $E(X|\mathcal{G})$.

**Proof.** Given a positive r.v. $X$, let $Q_X : \mathcal{F} \rightarrow \mathbb{R}$ be a set function satisfying

$$
Q_X(A) = \int_A X d\mathbb{P}, \quad (3.14)
$$

for $A \in \mathcal{F}$. It can be checked that $Q_X$ is a probability measure and $Q_X \ll \mathbb{P}$ on $\mathcal{G}$. Thus, applying Radon-Nikodym Theorem on the measure $Q_X(\cdot)$ yields a $\mathcal{G}$ measurable r.v., say $Y$, such that $Q_X(A) = \int_A Y d\mathbb{P}$ for $A \in \mathcal{G}$. Together with $(3.14)$, $Y$ satisfies i) and ii) of Definition (3.9). Thus $X_\mathcal{G} = Y$ is the desired conditional expectation of $X$ given $\mathcal{G}$.

For general r.v. $X$, we can write $X = X^+ - X^-$, where $X^+ = X1_{\{X > 0\}}$ and $X^- = -X1_{\{X < 0\}}$ are positive r.v.s. Then the preceding result can be applied to each of $X^+$ and $X^-$, and the difference of the two conditional expectations is the desired $X_\mathcal{G}$.

**Example 3.13.** Refer to Example 3.12, we have $\frac{d\mathbb{P}_1}{d\mathbb{P}_2}(\omega) = 3/2|\omega - 3.5|$, $\frac{d\mathbb{P}_3}{d\mathbb{P}_2}(\omega) = 2|\omega - 3.5|/3$, $\frac{d\mathbb{P}_5}{d\mathbb{P}_1}(\omega) = 2 \cdot 1_{\{\omega \geq 4\}}$ and $\frac{d\mathbb{P}_4}{d\mathbb{P}_1}(\omega) = 2 \cdot 1_{\{\omega \leq 3\}}$. What are $\frac{d\mathbb{P}_2}{d\mathbb{P}_1}$ and $\frac{d\mathbb{P}_3}{d\mathbb{P}_2}$?

On the other hand, note that $\mathbb{P}_N((\infty, x)) = F_N(x) = \int_x^\infty f_N(x) dx$ and $\mathbb{P}_E((\infty, x)) = F_E(x) = \int_x^\infty f_E(x) dx$ where $f_N(x)$ and $f_E(x)$ are the p.d.f.s of standard normal and standard exponential distribution, respectively. Note that $\frac{f_E(x)}{f_N(x)}$ is Borel measurable and satisfies

$$
\int_A \frac{f_E(x)}{f_N(x)} d\mathbb{P}_N = \int_A \frac{f_E(x)}{f_N(x)} f_N(x) dx = \int_A f_E(x) dx = \mathbb{P}_E(A).
$$

Thus, $\frac{d\mathbb{P}_E}{d\mathbb{P}_N} = \frac{f_E(x)}{f_N(x)}$. However, $\frac{d\mathbb{P}_N}{d\mathbb{P}_E}$ does not exist as $\mathbb{P}_N$ is not absolutely continuous w.r.t. $\mathbb{P}_E$ ($f_N(x)/f_E(x)$ does not exist for $x < 0$).

### 3.5 General Properties of Conditional Expectations

**Theorem 3.6.** Conditional expectation has the following properties

...
1. \(E(aX + bY|\mathcal{G}) = aE(X|\mathcal{G}) + bE(Y|\mathcal{G})\) (linearity).
2. \(E(E(X|\mathcal{G})) = E(X)\).
3. \(E(XY|\mathcal{G}) = XE(Y|\mathcal{G})\) if \(X\) is \(\mathcal{G}\)-measurable.
4. \(E(X|\mathcal{G}) = E(X)\) if \(X\) is independent of \(\mathcal{G}\)
5. \(E(E(X|\mathcal{G})|\mathcal{H}) = E(X|\mathcal{H})\) if \(\mathcal{H} \subset \mathcal{G}\) (tower property).
6. If \(X \geq 0\), then \(E(X|\mathcal{G}) \geq 0\) (positivity).

Here \(a, b\) are real numbers, \(X, Y \in L^1\) are defined on a probability space \((\Omega, \mathcal{F}, P)\) and \(\mathcal{G}, \mathcal{H} \subset \mathcal{F}\) are \(\sigma\)-fields. In (3), we also assume that the product \(XY \in L^1\). All equalities and inequalities hold almost surely.

Proof. 1. For any \(B \in \mathcal{G}\)
\[
\int_B (aE(X|\mathcal{G}) + bE(Y|\mathcal{G}))dP = a\int_B E(X|\mathcal{G})dP + b\int_B E(Y|\mathcal{G})dP
= \int_B (aX + bY)dP.
\]
The result thus follows from the definition of conditional expectation.

2. This follows by putting \(B = \Omega\) in Example 3.10. Also, it is a special case of (5) when \(\mathcal{H} = \{\emptyset, \Omega\}\).

3. We first verify the result for \(X = \mathbb{I}_A\) (indicator function) where \(A \in \mathcal{G}\). In this case
\[
\int_B \mathbb{I}_A E(Y|\mathcal{G})dP = \int_{A \cap B} E(Y|\mathcal{G})dP
= \int_{A \cap B} YdP
= \int_B \mathbb{I}_A YdP.
\]
for any \(B \in \mathcal{G}\). From the definition of conditional expectation, this implies
\[
\mathbb{I}_A E(Y|\mathcal{G}) = E(\mathbb{I}_A Y|\mathcal{G})\).
\]
Using 1), we can extend the preceding result to a simple-function \(X = \sum_{j=1}^m a_j \mathbb{I}_{A_j}\), i.e.,
\[
XE(Y|\mathcal{G}) = \sum_{j=1}^m a_j \mathbb{I}_{A_j} E(Y|\mathcal{G}) = E(\sum_{j=1}^m a_j \mathbb{I}_{A_j} Y|\mathcal{G}) = E(XY|\mathcal{G})\).
\]
where \(A_j \in \mathcal{G}\) for \(j = 1, 2, \ldots, m\). Finally, the general result follows by approximating \(X\) with an increasing sequence of simple functions and use MCT.

4. Since \(X\) is independent of \(\mathcal{G}\), the random variables \(X\) and \(\mathbb{I}_B\) are independent for any \(B \in \mathcal{G}\). Thus
\[
\int_B E(X)dP = E(X) \int_B dP = E(X) \int_{\Omega} \mathbb{I}_B dP = E(X)E(\mathbb{I}_B).
\]
3.5 General Properties of Conditional Expectations

Since independent random variables are uncorrelated,

$$E(X)E(Y) = E(XY) = \int_B XdP.$$ 

Now we have obtained that

$$\int_B E(X)dP = \int_B XdP$$

for all $B \in \mathcal{G}$, i.e., $E(X|\mathcal{G}) = E(X)$.

5. By Definition,

$$\int_B E(X|\mathcal{G})dP = \int_B XdP$$

for every $B \in \mathcal{G}$ and

$$\int_B E(X|\mathcal{H})dP = \int_B XdP$$

for every $B \in \mathcal{H}$. Since $\mathcal{H} \subset \mathcal{G}$,

$$\int_B E(X|\mathcal{G})dP = \int_B E(X|\mathcal{H})dP$$

for every $B \in \mathcal{H}$. Therefore, the definition of conditional expectation implies that

$$E(E(X|\mathcal{G})|\mathcal{H}) = E(X|\mathcal{H}).$$

6. For any $n$ we put

$$A_n = \left\{ E(X|\mathcal{G}) \leq -\frac{1}{n} \right\}.$$

Since $E(X|\mathcal{G})$ is a $\mathcal{G}$ measurable r.v., we have $A_n \in \mathcal{G}$. If $X \geq 0$ a.s., then

$$0 \leq \int_{A_n} XdP = \int_{A_n} E(X|\mathcal{G})dP \leq -\frac{1}{n}P(A_n)$$

which implies $P(A_n) = 0$. Since

$$\{E(X|\mathcal{G}) < 0\} = \bigcup_{n=1}^{\infty} A_n,$$

it follows that $P\{E(X|\mathcal{G}) < 0\} = \sum_{n=1}^{\infty} P(A_n) = 0$, completing the proof.

Example 3.14. (Independence and Zero-Correlation) Recall that $X \in \mathcal{L}^1$ if $E(|X|) < \infty$. If $X_1$ and $X_2$ are independent, then they are uncorrelated, i.e., $E(X_1X_2) = E(X_1)E(X_2)$. To see this, note that

$$E(X_1X_2) = E(E(X_1X_2|\sigma(X_2))) = E(E(X_2E(X_1|\sigma(X_2)))) = E(X_2E(X_1)) = E(X_1)E(X_2),$$

(3.15)

where the four equalities are consequences of Properties 2., 3., 4., 1. of Theorem 3.6 respectively. Repeating the argument in (3.15), we have that, if $X_1, \ldots, X_n \in \mathcal{L}^1$ are independent, then they are uncorrelated, i.e.,
\[ E(X_1X_2 \cdots X_n) = E(X_1)E(X_2) \cdots E(X_n) \]

provided that the product \(X_1X_2 \cdots X_n \in L^1\). However, the opposite direction is not true. See Exercise 3.15.

### 3.6 Useful Inequalities

**Definition 3.11. (Convex Function)** A function \( \varphi : \mathbb{R} \to \mathbb{R} \) is convex if for any \( x, y \in \mathbb{R} \) and any \( \lambda \in [0, 1] \)

\[
\varphi(\lambda x + (1 - \lambda)y) \leq \lambda \varphi(x) + (1 - \lambda)\varphi(y).
\]

Graphically, a function \( \varphi \) is convex if it lies below the straight line from the point \((x, \varphi(x))\) to the point \((y, \varphi(y))\) on any interval \([x, y]\).

**Theorem 3.7. (Jensen’s Inequality)** Let \( \varphi : \mathbb{R} \to \mathbb{R} \) be a convex function and let \( X \in L^1 \) on \((\Omega, \mathcal{F}, P)\) such that \( \varphi(X) \in L^1 \). Then

\[
\varphi(E(X|\mathcal{G})) \leq E(\varphi(X)|\mathcal{G}) \quad \text{a.s.} \tag{3.16}
\]

for any \( \sigma \)-field \( \mathcal{G} \) on \( \Omega \) contained in \( \mathcal{F} \).

**Proof.** By the convexity of \( \varphi \), for every fixed \( c \in \mathbb{R} \), there exists an \( a_c \) such that \( \varphi(x) \geq a_c(x - c) + \varphi(c) \) for all \( x \). (If \( \varphi(x) \) is continuous and \( a_c = \varphi'(x)|_{x=c} \), then the line \( a_c(x - c) + \varphi(c) \) is the tangent of \( \varphi(x) \) at the point \((c, \varphi(c))\).) Replacing \( x \) by a r.v. \( X \) and taking conditional expectation on both sides gives

\[
E(\varphi(X)|\mathcal{G}) \geq a_c(E(X|\mathcal{G}) - c) + \varphi(c).
\]

Since \( c \) is arbitrary, put \( c = E(X|\mathcal{G}) \) yields (3.16). \( \square \)

**Theorem 3.8. (Hölder Inequality)** Let \( p > 1 \), \( \frac{1}{p} + \frac{1}{q} = 1 \) and \( X, Y \) be r.v. on \((\Omega, \mathcal{F}, P)\). If \( E(X^p) < \infty \) and \( E(Y^q) < \infty \), then \( E(|XY|) < \infty \) and

\[
E(XY) \leq (E(X^p))^{\frac{1}{p}}(E(Y^q))^{\frac{1}{q}}.
\]

**Proof.** Without loss of generality, assume that \( X \) and \( Y \) are positive r.v.s. First define a new probability measure \( Q \) on \((\Omega, \mathcal{F})\) by

\[
Q(A) = \int_A \frac{X^p}{||X||_p^p} \, d\mathbb{P},
\]

where \( ||X||_p = (E(X^p))^{1/p} \). Then,
Theorem 3.10. We have the following generalization of MCT, Fatou’s lemma and DCT:

i) (Monotone Convergence Theorem (MCT)) Let $\Omega$, $\mathcal{F}$, $P$ be a sub $\sigma$-field of $\mathcal{F}$.

ii) (Fatou’s Lemma) If $\Omega$, $\mathcal{F}$, $P$ be a sub $\sigma$-field of $\mathcal{F}$, then

iii) (Dominated Convergence Theorem (DCT)) If for some r.v. $M$ and all $n \geq 1$, $X_n \leq M$ (i.e. $|X_n(\omega)| \leq M(\omega)$ for all $\omega \in \Omega$), and if $E(M) \leq \infty$ and $X_n \overset{a.s.}{\to} X$, then

$$E(X_n|\mathcal{G}) \overset{a.s.}{\to} E(X|\mathcal{G}).$$
3.7 Exercises

Exercise 3.11 Consider the dice example \((\Omega, \mathcal{F}, P)\) where \(\Omega = \{1, 2, 3, 4, 5, 6\}\) and \(\mathcal{F} = 2^\Omega\). Let \(P(\{\omega\}) = 1/6\) for \(\omega = 1, \ldots, 6\), and

\[ A_1 = \{1, 2, 3\}, A_2 = \{4, 5, 6\}, A_3 = \{2, 5\}, A_4 = \{2, 6\}, \]

\[ X_i = 1_{A_i}, \mathcal{F}_i = \sigma(\{A_i\}) \text{ for } i = 1, \ldots, 4, \]

\[ X_5(\omega) = \omega 1_{A_1}(\omega), \mathcal{F}_5 = \sigma(\{A_1, A_3\}), \mathcal{F}_6 = \sigma(\{A_3, A_4\}), \mathcal{F}_7 = 2^\Omega. \]

a) Which events are independent? 
b) Which random variables are independent? 
c) Which \(\sigma\)-fields are independent? 
d) Repeat a)-c) using the measure \(P(\{\omega\}) = |\omega - 3.5|/9\) for \(\omega = 1, \ldots, 6\).

Exercise 3.12 Explain the meaning of independence between two events and mutually exclusive between two events. Can two events be both independent and mutually exclusive? Explain.

Exercise 3.13 How to define the independence between an event and a random variable? How to define the independence between an event and a \(\sigma\)-field? How to define the independence between a random variable and a \(\sigma\)-field?

Exercise 3.14 Show that two random variables \(X\) and \(Y\) are independent if and only if the \(\sigma\)-fields \(\sigma(X)\) and \(\sigma(Y)\) are independent.

Exercise 3.15 Let \(X \sim N(0, 1)\) and \(Y = X^2\). Show that \(X\) and \(Y\) are uncorrelated but dependent.

Exercise 3.16 Let \(I_A\) be the indicator function of an event \(A\), i.e.

\[ I_A(\omega) = \begin{cases} 1 & \text{if } \omega \in A \\ 0 & \text{if } \omega \notin A. \end{cases} \]

Show that \(E(I_A|B) = P(A|B)\) for any event \(B\) with \(P(B) \neq 0\).

Exercise 3.17 For the following r.v.s, write down the probability space and the induced probability space, and find the expectation of \(X\).

i) (Constant r.v.) \(X(\omega) = a\) for all \(\omega\). 
ii) \(X : [0, 1] \to \mathbb{R}\) given by \(X(\omega) = \min\{\omega, 1 - \omega\}\) (The distance to the nearest endpoint of the interval \([0, 1]\))
iii) \(X : [0, 1]^2 \to \mathbb{R}\), the distance to the nearest endpoint of the unit square.

Exercise 3.18 Let \(\Omega = [0, 1]\), \(\mathcal{F} = \mathcal{B}_{[0, 1]}\) and \(P = \lambda_{[0,1]}\). Let \(X(\omega) = \omega(1 - \omega)\) for \(\omega \in \Omega\). For any random variable \(Y\) on \((\Omega, \mathcal{F}, P)\),

i) Show that \(Y(\omega) + Y(1 - \omega)\) is \(\sigma(X)\) measurable. (Hint: Draw the graph \(Y\) against \(\omega\))
ii) Show that $E(Y|X)(\omega) = \frac{Y(\omega)+Y(\omega)}{2}$ for any $\omega \in \Omega$.

**Exercise 3.19** Take $\Omega = [0,1]$, $G = [0,1]$ and $\mathbb{P} = \lambda_{[0,1]}$, the Lebesgue measure on $[0,1]$. Find $E(X|Z)$ if $X(\omega) = 2\omega^2$ and $Z(\omega) = 1_{\{\omega \in [0,1/3]\}} + 2 \times 1_{\{\omega \in (2,3,1]\}}$.

**Exercise 3.20** Show that if $Z$ is a constant function, then $E(X|Z)$ is constant and equal to $E(X)$.

**Exercise 3.21** On $(\Omega, \mathcal{F}, \mathbb{P})$, show that for $A, B \in \mathcal{F}$,

$$E(1_A|1_B)(\omega) = \begin{cases} \mathbb{P}(A|B) & \text{if } \omega \in B, \\ \mathbb{P}(A|\Omega \setminus B) & \text{if } \omega \notin B. \end{cases}$$

**Exercise 3.22** Recall that if $Z : \Omega \rightarrow \mathbb{R}$ is a discrete r.v. taking $n$ values $\{z_i\}_{i=1,...,n}$, then there exists disjoint $A_i$ such that $\Omega = \bigcup_{i=1}^n A_i$ and $Z(\omega) = z_i$ whenever $\omega \in A_i$. Show that every element in $\sigma(Z)$ is a union of some $A_i$. Generalize to the case where $Z$ is a discrete r.v. taking infinite values $\{z_i\}_{i=1,...}$. 

**Exercise 3.23** Show that if $X$ is $\mathcal{G}$-measurable, then $E(X|\mathcal{G}) = X$ a.s.

**Exercise 3.24** Refer to Example 3.10. If $B \notin \mathcal{G}$, does the result still hold? Prove it or give a counter example.

**Exercise 3.25** Show that, if $\mathbb{Q} \ll \mathbb{P}$, then for any $\epsilon > 0$, we can find a $\tilde{\delta} > 0$ such that $\mathbb{P}(F) < \tilde{\delta}$ implies $\mathbb{Q}(F) \leq \epsilon$.

**Exercise 3.26** Show that, if $\mathbb{P}$ and $\mathbb{Q}$ satisfy (3.11) for any $A \in \mathcal{F}$, then $\mathbb{Q} \ll \mathbb{P}$.

**Exercise 3.27** Let $\mathbb{Q}_1$, $\mathbb{Q}_2$ and $\mathbb{P}$ be measures on $(\Omega, \mathcal{F})$. Show that if $\mathbb{Q}_1 \ll \mathbb{P}$ and $\mathbb{Q}_2 \ll \mathbb{P}$, then $\mathbb{Q}_1 + \mathbb{Q}_2 \ll \mathbb{P}$.

**Exercise 3.28** Consider the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with $\Omega = \{1,2,3,4,5,6\}$ and $\mathcal{F} = 2^\Omega$. If $\mathbb{P}(\{i\}) = i/21$ and $\mathbb{Q}(\{i\}) = (i-1)/15$. Is $\mathbb{P} \ll \mathbb{Q}$? Is $\mathbb{Q} \ll \mathbb{P}$? Find the Radon-Nikodym derivative $d\mathbb{P}/d\mathbb{Q}$ and $d\mathbb{Q}/d\mathbb{P}$, if they exist.

**Exercise 3.29** Justify rigorously the equality $\mathbb{P}(\bigcap_{n=1}^\infty A_n) = \lim_{n \rightarrow \infty} \mathbb{P}(\bigcap_{n=1}^m A_n)$ in (3.5).

**Exercise 3.30** Using the notation in the proof of Theorem 3.4.

i) From the existence of $l = \lim_{k \rightarrow \infty} \int_\Omega Y_k^2 \ d\mathbb{P}$, show that we can find a subsequence $\{n_k\}_{k=1,...}$ such that $l - \frac{1}{\mathbb{P}} < \lim_{k \rightarrow \infty} \int_\Omega Y_{n_k}^2 \ d\mathbb{P} \leq l$.

ii) Using the definition of $Y_k$, show that $\int_\Omega (Y_{n_k+1} - Y_{n_k})^2 \ d\mathbb{P} = \int_\Omega (Y_{n_k+1}^2 - Y_{n_k}^2) \ d\mathbb{P} < \frac{1}{\mathbb{P}}$.

iii) By Cauchy Schwartz inequality, show that $\int_\Omega |Y_{n_k+1} - Y_{n_k}| \ d\mathbb{P} < \frac{1}{\mathbb{P}}$.

iv) Using (iii) and MCT, show that $\int_\Omega \sum_{k=1}^\infty |Y_{n_k+1} - Y_{n_k}| \ d\mathbb{P} = \sum_{k=1}^\infty \int_\Omega |Y_{n_k+1} - Y_{n_k}| \ d\mathbb{P} < \infty$.

v) Using (iv), argue that $\sum_{k=1}^\infty |Y_{n_k+1} - Y_{n_k}| < \infty$ almost surely. (Hint: proof by contradiction). Then, show that $\sum_{k=1}^\infty (Y_{n_k+1} - Y_{n_k}) < \infty$ almost surely.
vi) Using (v), show that \( \lim_{k \to \infty} Y_k \) exists almost surely.

vii) Noting that \( l - \frac{1}{4k} < \int_Y Y^2_n \leq \int_Y Y^2_{\mathcal{F}_k} \leq l \), repeating the arguments (ii)-(iii), show that

\[
\left| \int_{\mathcal{F}} (Y^2_{\mathcal{F}_k} - Y^2_n) \, dP \right| < 1/2^k.
\]

viii) Using (vi) and (vii), show that \( \int_{\mathcal{F}} Y_{\mathcal{F}_k} \, dP \to \int_{\mathcal{F}} Y \, dP \).

ix) Using \( \int_A Y \, dP = P(A) = \int_A dP \), show that \( E_{\mathcal{F}}(YY_F) = E_P(Y) \) for any simple function \( Y \) on \( \mathcal{F} \). Extend to the case where \( Y \) is \( \mathcal{F} \)-measurable.