The Monotone Convergence Theorem

Here we are going to describe, illustrate, and prove a famous and important theorem from measure theory as applied to discrete random variables.

**Theorem.** Let $X$ be a discrete random variable and let $\{X_n\}_{n=1}^{\infty}$ be a sequence of discrete random variables such that, for almost every $\omega$,

(i) $0 \leq X_1(\omega) \leq X_2(\omega) \leq X_3(\omega) \leq \ldots \leq X_n(\omega) \leq \ldots \leq X(\omega)$ and  
(ii) $\lim_{n \to \infty} X_n(\omega) = X(\omega)$.

Then $\lim_{n \to \infty} E[X_n] = E[X]$.

**Note:** The phrase “for almost every $\omega$” means that (i) and (ii) hold for all $\omega$ except for possibly on a set having probability 0. The phrase is often stated as “almost surely” abbreviated “a.s.”, or “almost everywhere” abbreviated “a.e.”.

To prove the theorem, we shall need several background results.

**Lemma 1.** Let $Y$ be a discrete random variable such that $Y \geq 0$ a.s. Then $E[Y] \geq 0$.

**Proof.** Let $\{y_i\}$ be all the non-negative values in the range of $Y$. Any negative values occur with probability 0 and therefore will not affect the mean. Thus,

$$E[Y] = \sum_{i} y_i \times P(Y = y_i) \geq 0.$$

**Corollary 1.** Let $Y$ and $W$ be a discrete random variables such that $W \leq Y$ a.s. Then $E[W] \leq E[Y]$.

**Proof.** The random variable $Y - W$ is still discrete and $Y - W \geq 0$ a.s. Thus, by Lemma 1, we have $0 \leq E[Y - W] = E[Y] - E[W]$.

**Lemma 2.** Given a nested increasing sequence of events $A_1 \subseteq A_2 \subseteq A_3 \subseteq \ldots \subseteq A_n \subseteq \ldots$ we have $\lim_{n \to \infty} P(A_n) = P(\bigcup_{i=1}^{\infty} A_i)$.

**Proof.** We first disjointify the events as follows: Let $B_1 = A_1$, $B_2 = A_2 - A_1$, $\ldots$, $B_n = A_n - A_{n-1}$. Then $\bigcup_{i=1}^{n} B_i = \bigcup_{i=1}^{n} A_i = A_n$, for all $n$, and $\bigcup_{i=1}^{\infty} B_i = \bigcup_{i=1}^{\infty} A_i$. Hence,

$$\lim_{n \to \infty} P(A_n) = \lim_{n \to \infty} P\left(\bigcup_{i=1}^{n} B_i\right) = \lim_{n \to \infty} \sum_{i=1}^{n} P(B_i) = \sum_{i=1}^{\infty} P(B_i) = P\left(\bigcup_{i=1}^{\infty} B_i\right) = P\left(\bigcup_{i=1}^{\infty} A_i\right).$$
**Corollary 2.** Given a nested decreasing sequence of events \( A_1 \supseteq A_2 \supseteq A_3 \supseteq \ldots \supseteq A_n \supseteq \ldots \) we have \( \lim_{n \to \infty} P(A_n) = P \left( \bigcap_{i=1}^{\infty} A_i \right) \).

**Proof.** The complements of the events are nested increasing: \((A_1)^c \subseteq (A_2)^c \subseteq (A_3)^c \subseteq \ldots\) Thus, by Lemma 2 and DeMorgan’s Law, we have

\[
\lim_{n \to \infty} P(A_n) = \lim_{n \to \infty} 1 - P((A_n)^c) = 1 - \lim_{n \to \infty} P((A_n)^c) = 1 - P \left( \bigcup_{i=1}^{\infty} (A_i)^c \right)
= 1 - P \left( \bigcap_{i=1}^{\infty} A_i \right) = P \left( \bigcap_{i=1}^{\infty} (A_i)^c \right).
\]

**Lemma 3.** Given random variables such that \( W \leq Y \) a.s., then a.s.,

\[
\{ \omega : W(\omega) \leq t \} \supseteq \{ \omega : Y(\omega) \leq t \}, \text{ for any given } t.
\]

**Proof.** Let \( \omega \) be in the set of probability 1 where \( W \leq Y \). If \( Y(\omega) \leq t \), then \( W(\omega) \leq t \) also because \( W(\omega) \leq Y(\omega) \leq t \); hence, we have the required set containment.

The final preliminary result below follows immediately from Lemma 3 and Corollary 2:

**Corollary 3.** If \( 0 \leq X_1(\omega) \leq X_2(\omega) \leq X_3(\omega) \leq \ldots \leq X_n(\omega) \leq \ldots \) a.s., then for all \( t \),

\[
\lim_{n \to \infty} P(X_n \leq t) = P \left( \bigcap_{i=1}^{\infty} \{ X_i \leq t \} \right).
\]

To illustrate the Monotone Convergence Theorem, consider a population of three students \( \Omega = \{ \omega_1, \omega_2, \omega_3 \} \), and let \( X_i \) be the number of hours earned after the students’ \( i \)th semester, and let \( X \) be the total number of hours finally earned. The chart that follows gives an example of the conditions of the theorem. Notice that

(i) Each \( X_i \) has a different range, and the values are not necessarily integers.

(ii) For each \( \omega \), \( 0 \leq X_1(\omega) \leq X_2(\omega) \leq X_3(\omega) \leq \ldots \leq X_n(\omega) \leq \ldots \leq X(\omega) \)

(iii) \( E[X_1] \leq E[X_2] \leq E[X_3] \leq \ldots \leq E[X] \)

(iv) \( P(X_1 \leq t) \geq P(X_2 \leq t) \geq P(X_3 \leq t) \geq P(X_4 \leq t) \geq \ldots \geq \ldots \)
Proof of MCT – Case I: Assume $E[X] < \infty$. Let $\epsilon > 0$ be given. Because $X \geq 0$ a.s., we only need to sum over the positive values $\{k_i\}$ in its range in order to compute its mean. We shall assume that $0 < k_1 < k_2 < k_3 < \ldots$.

Now because $E[X] = \sum_{i=1}^{\infty} k_i P(X = k_i) = \lim_{n \to \infty} \sum_{i=1}^{n} k_i P(X = k_i)$, there exists an integer $N \geq 1$ such that $\sum_{i=1}^{N} k_i P(X = k_i) > E[X] - \frac{\epsilon}{3}$. That is, $\sum_{i=N+1}^{\infty} k_i P(X = k_i) < \frac{\epsilon}{3}$. Because

$$\sum_{i=N+1}^{\infty} P(X = k_i) \leq \sum_{i=N+1}^{\infty} k_i P(X = k_i) < \frac{\epsilon}{3},$$

virtually all of the weight and all of the weighted average of $X$ occurs for $1 \leq X \leq k_N$. Then because each $X_n \leq X$, virtually all their weights and weighted averages will also occur for $1 \leq X_n \leq k_N$.

By Corollary 3, for any fixed $t$, the sequence $\{P(X_n \leq t)\}_n$ converges to $P\left(\bigcap_{i=1}^{\infty} \{X_i \leq t\}\right)$. But what is this intersection? If $X_i(\omega) \leq t$ for all $i$, then $X(\omega) = \lim_{i \to \infty} X_i(\omega) \leq t$. And if $X(\omega) \leq t$, then $X_i(\omega) \leq X(\omega) \leq t$ for all $i$. Hence, a.s., the intersection is $\{X \leq t\}$. Therefore, $\lim_{n \to \infty} P(X_n \leq t) = P\left(\bigcap_{i=1}^{\infty} \{X_i \leq t\}\right) = P(X \leq t)$.

Because there is separation between the $k_i$ ($k_1 < k_2 < k_3 < \ldots < k_N$), we can choose a $\delta > 0$ with $0 < \delta < \epsilon / 3$, such that $k_{i-1} < k_i - \delta < k_i$, for $2 \leq i \leq N$, forming disjoint intervals

$$A_1 = (k_1 - \delta, k_1] \quad A_2 = (k_2 - \delta, k_2] \quad \ldots, \quad A_N = (k_N - \delta, k_N]$$
For each of these finite number of intervals, we have

\[ \lim_{n \to \infty} P(X_n \in A_i) = \lim_{n \to \infty} (P(X_n \leq k_i) - P(X_n \leq k_i - \delta)) = P(X \leq k_i) - P(X \leq k_i - \delta) = P(X = k_i). \]

So there exists an integer \( M \geq 1 \) such that if \( n \geq M \) then

\[ |P(X_n \in A_i) - P(X = k_i)| < \frac{\varepsilon}{3NkN} \quad \text{for all } i = 1, \ldots, N. \]

Then for all \( n \geq M \),

\[
\begin{align*}
E[X_n] &\geq \sum_{i=1}^{N} E[X_n \mid X_n \in A_i] \times P(X_n \in A_i) \\
&\geq \sum_{i=1}^{N} (k_i - \delta) \times P(X_n \in A_i) \\
&\geq \sum_{i=1}^{N} k_i \left( P(X = k_i) - \frac{\varepsilon}{3NkN} \right) - \delta \\
&= \sum_{i=1}^{N} k_i \left( P(X = k_i) \right) - \frac{\varepsilon}{3NkN} \sum_{i=1}^{N} k_i - \delta \\
&\geq \sum_{i=1}^{N} k_i \left( P(X = k_i) \right) - \frac{\varepsilon N k_N}{3NkN} - \delta \\
&> E[X] - \frac{\varepsilon}{3} - \frac{\varepsilon}{3} - \frac{\varepsilon}{3} \\
&= E[X] - \varepsilon.
\end{align*}
\]

Thus for all \( n \geq M \), we have \( E[X] - \varepsilon < E[X_n] \leq E[X] \), which proves that

\[ \lim_{n \to \infty} E[X_n] = E[X]. \]

**Proof of MCT – Case II:** Assume \( E[X] = \infty \). Let \( L > 0 \) be given. Then there exists an integer \( M \geq 1 \) such that \( \sum_{i=1}^{N} k_i P(X = k_i) > L + 1 \). We now choose a \( \delta > 0 \) with \( 0 < \delta < 1/2 \), such that \( k_{i-1} < k_i - \delta < k_i \), for \( 2 \leq i \leq N \), forming the disjoint intervals

\[ A_1 = (k_1 - \delta, k_1], \quad A_2 = (k_2 - \delta, k_2], \ldots, \quad A_N = (k_N - \delta, k_N] \]

Again we have \( \lim_{n \to \infty} P(X_n \in A_i) = P(X = k_i) \) for each interval, so there exists an integer \( M \geq 1 \) such that if \( n \geq M \) then

\[ |P(X_n \in A_i) - P(X = k_i)| < \frac{1}{2NkN} \quad \text{for all } i = 1, \ldots, N. \]
Then for all \( n \geq M \),

\[
E[X_n] \geq \sum_{i=1}^{N} E[X_n \mid X_n \in A_i] \times P(X_n \in A_i) \geq \sum_{i=1}^{N} (k_i - \delta) \times P(X_n \in A_i)
\]

\[
= \sum_{i=1}^{N} k_i P(X_n \in A_i) - \delta \sum_{i=1}^{N} P(X_n \in A_i)
\]

\[
\geq \sum_{i=1}^{N} k_i \left( P(X = k_i) - \frac{1}{2N k_N} \right) - \delta
\]

\[
= \sum_{i=1}^{N} k_i P(X = k_i) - \frac{N k_N}{2N k_N} \sum_{i=1}^{N} k_i - \delta
\]

\[
\geq \sum_{i=1}^{N} k_i P(X = k_i) - \frac{N k_N}{2N k_N} \approx L + 1 - \frac{1}{2} - \frac{1}{2}
\]

\[
= L.
\]

Thus \( \lim_{n \to \infty} E[X_n] = \infty \).