Martingale Convergence Theorem

Auriel Rosenzweig

Abstract
We will define discrete-time Martingales, and prove two Martingale convergence theorems, as well as discuss several of their applications in Measure Theory, Ergodic Theory and Probability Theory. The talk will be roughly based on the book "Ergodic theory with a view toward Number Theory", by Einsiedler and Ward.

1 Basic Definitions and Examples

Definition 1.1 (Filtration). Let $(X, B, \mu)$ be a measure space. A filtration is a sequence of sub-$\sigma$-algebras of $B$, such that $F_1 \subseteq F_2 \subseteq \ldots \subseteq B$. We define:

$$F_\infty := \sigma \left( \bigcup_{n \in \mathbb{N}} F_n \right) \subseteq B.$$ 

We will call $(X, B, \mu, (F_n)_{n \in \mathbb{N}})$ a filtered space.

Definition 1.2 (Martingale). Let $(X, B, \mu, (F_n)_{n \in \mathbb{N}})$ be a filtered space. A Martingale is a process $M = (M_n)_{n \in \mathbb{N}}$ relative to the filtration $(F_n)_{n \in \mathbb{N}}$ and the measure $\mu$, which satisfies:

1. $M_n \in L^1(F_n)$, for every $n \in \mathbb{N}$. I.e., $M_n$ is integrable, and $F_n$ measurable.
2. $E[M_{n+1} | F_n] = M_n$, for every $n \in \mathbb{N}$.

The next family of examples will be the general case for both of the theorems we will prove today.

Example 1.3. Let $(X, B, \mu, (F_n)_{n \in \mathbb{N}})$ be a filtered space. Let $f \in L^1(B)$. Define $M_n = E[f | F_n]$. Note that $(M_n)$ is a Martingale relative to $(F_n)$. Indeed:

- By the definition of the conditional expectation, the process $M_n$ is adapted to $(F_n)$, and $M_n \in L^1$.
- By the tower property,

$$E[M_{n+1} | F_n] = E[E[f | F_{n+1}] | F_n] = [f | F_n] = M_n.$$ 

The following theorem, is the most general Martingale convergence theorem.
**Theorem 1.4** (Doob's Martingale Convergence Theorem). Let \((M_n)\) be a Martingale relative to \((\mathcal{F}_n)\), and suppose that \(M_n\) is uniformly bounded in \(L^1\), i.e.,

\[
\sup_n \|M_n\| < \infty.
\]

There exists a random variable \(M_\infty \in L^1(\mathcal{F}_\infty)\), such that:

\[
M_n \xrightarrow{\text{a.s.}} M_\infty.
\]

We won’t prove Theorem 1.4 in this talk, but rather a theorem that holds only in the setting of Example 1.3. The theorem that we prove, will give us also convergence in \(L^1\) in addition to the almost surely convergence we get by Theorem 1.4.

The next example will show us that in the settings of Theorem 1.4, we can’t assure convergence in \(L^1\).

**Example 1.5.** Let \(X = [0, 1)\) and let \(\mu\) be the Lebesgue measure. Define the filtration

\[
\mathcal{F}_n = \sigma\left(\left\{\frac{1}{k+1}, \frac{1}{k} \mid 1 \leq k \leq n\right\} \cup \left\{[0, \frac{1}{n+1})\right\}\right),
\]

and the process \(M_n = (n+1) \mathbf{1}_{[0, \frac{1}{n+1})}\), for every \(n \in \mathbb{N}\).

Direct computation will show that \(M_n\) is a Martingale, and that \(\mathbb{E}[M_n] = 1\) for all \(n \in \mathbb{N}\), and therefore \(M_n\) is uniformly bounded in \(L^1\). Moreover, it is clear that \(M_n \xrightarrow{\text{a.s.}} 0\). However,

\[
\lim_{n \to \infty} \mathbb{E}[M_n] = 1 \neq 0 = \mathbb{E}[0],
\]

and therefore \(M_n\) does not converge to \(M_\infty\) in \(L^1\).

## 2 Increasing Martingale Theorem

In this section we prove the increasing Martingale theorem which sometimes referred as the Levy’s ‘Upward’ Theorem.

**Theorem 2.1.** Let \((X, \mathcal{B}, \mu, (\mathcal{F}_n)_{n \in \mathbb{N}})\) be a filtered space. For every \(f \in L^1(\mathcal{B})\),

\[
\mathbb{E}[f \mid \mathcal{F}_n] \xrightarrow{n \to \infty} \mathbb{E}[f \mid \mathcal{F}_\infty],
\]

almost everywhere and in \(L^1\).

In order to prove Theorem 2.1, we will use the next Lemma.

**Lemma 2.2** (Doob’s inequality). Let \(f \in L^1(\mathcal{B})\), and let \((\mathcal{F}_n)_{n \in \mathbb{N}}\) be a filtration. Fix \(\lambda > 0\) and let

\[
E = \left\{ x \in X \mid \sup_{n \geq 1} \mathbb{E}[f \mid \mathcal{F}_n] > \lambda \right\},
\]

Then

\[
\mu(E) \leq \frac{1}{\lambda} \|f\|_1.
\]
Proof. We may assume without loss of generality that $f \geq 0$. Otherwise, we prove it for $|f|$ which can only make $\mu(E)$ larger. Let

$$E^N = \left\{ x \in X \mid \max_{1 \leq i \leq N} E[f \mid F_i](x) > \lambda \right\},$$

for every $N \in \mathbb{N}$, and for every $1 \leq n \leq N$, let

$$E^N_n = \left\{ x \in X \mid E[f \mid F_n](x) > \lambda \text{ and } E[f \mid F_i](x) \leq \lambda \text{ for } 1 \leq i \leq n-1 \right\}.$$

Note that $E^N = \bigcup_{n=1}^N E^N_n$, and $E^N_n \in F_n$ since $F_1, F_2, \ldots, F_{n-1} \subseteq F_n$. Thus,

$$\lambda \mu(E^N) = \sum_{n=1}^N \lambda \mu(E^N_n) = \sum_{n=1}^N \int_{E^N_n} \lambda \leq \sum_{n=1}^N \int_{E^N_n} E[f \mid F_n] = \sum_{n=1}^N \int_{E^N_n} f = \int_{E^N} f \leq \|f\|_1.$$

By measure continuity, we deduce

$$\mu(E) = \mu\left( \bigcup_{N \in \mathbb{N}} E^N \right) = \lim_{N \to \infty} \mu(E^N) \leq \frac{1}{\lambda} \|f\|_1.$$

Lemma 2.3. Let $(X, \mathcal{B}, \mu, (F_n)_{n \in \mathbb{N}})$ be a filtered space. Then $\bigcup_{n \in \mathbb{N}} L^1(F_n)$ is dense in $L^1(F_\infty)$.

The proof of the lemma is left as an exercise. Prove the density for an indicator function, then simple function and then any $L^1$ function. The first part can be proved by showing that the collection

$$\{ A \in \mathcal{B} \mid \forall \varepsilon > 0 \exists (m \in \mathbb{N} \land A_m \in F_m) \text{ such that } \mu(A \triangle A_m) < \varepsilon \},$$

is a $\sigma$-algebra that contains $\bigcup_{n \in \mathbb{N}} F_n$ and thus contains $F_\infty$.

Proof of Theorem 2.1. Since $E[E[f \mid F_\infty] \mid F_n] = E[f \mid F_n]$, we may assume without loss of generality that $f$ is $F_\infty$ measurable.

If there exists $m \in \mathbb{N}$ such that $f$ is $F_m$ measurable, then $[f \mid F_n]$ is eventually constant sequence and thus the theorem holds. From Lemma 2.3, for every $\varepsilon > 0$, there exists $m \in \mathbb{N}$ and $g \in L^1(F_m)$ such that $\|f - g\|_1 < \frac{\varepsilon}{2}$. Thus, for every $n \geq m$,

$$\|E[f \mid F_n] - f\|_1 \leq \|E[f \mid F_n] - E[g \mid F_n]\|_1 + \|f - g\|_1 + \|E[g \mid F_n] - g\|_1 < \varepsilon.$$
For the almost everywhere convergence, we note that

\[
\mu \left( \left\{ x \in X \mid \limsup_{n \to \infty} |E[f \mid F_n] - f(x)| > \sqrt{\varepsilon} \right\} \right) \\
= \mu \left( \left\{ x \in X \mid \limsup_{n \to \infty} |E[f - g \mid F_n](x) - (f - g)(x)| > \sqrt{\varepsilon} \right\} \right) \\
\leq \mu \left( \left\{ x \in X \mid \sup_{n \in \mathbb{N}} |E[|f - g| \mid F_n](x)| > \frac{\sqrt{\varepsilon}}{2} \right\} \right) + \mu \left( \left\{ x \in X \mid |f - g|(x) > \frac{\sqrt{\varepsilon}}{2} \right\} \right)
\]

Lemma 2.2 \[2 \sqrt{\varepsilon} \|f - g\|_1 + 2 \sqrt{\varepsilon} \|f - g\|_1 \leq 2 \sqrt{\varepsilon}.
\]

Since \( \varepsilon \) is arbitrary, it follows that \( \limsup_{n \to \infty} |E[f \mid F_n] - f| = 0 \) almost everywhere. \( \square \)

### 3 Decreasing Martingale Theorem

In this section we prove the decreasing Martingale theorem which sometimes referred as the Levy's 'Downward' Theorem.

**Definition 3.1 (Backward Martingale).** Let \((X, \mathcal{B}, \mu)\) be a probability space. Let \((F_{-n})_{n \in \mathbb{N}}\) be a decreasing sequence of \(\sigma\)-algebras. I.e.,

\[B \supseteq F_{-1} \supseteq F_{-2} \supseteq \ldots \supseteq \bigcap_{n \in \mathbb{N}} F_{-n} := F_{-\infty}.
\]

A process \((M_{-n})_{n \in \mathbb{N}}\) is said to be backward Martingale, if:

1. \(M_n \in L^1(B_{-n})\), for every \(n \in \mathbb{N}\).
2. \(E[M_{-n+1} \mid F_{-n}] = M_{-n}\), for every \(n \in \mathbb{N}\).

**Theorem 3.2.** Let \((X, \mathcal{B}, \mu)\) be a probability space, and let \((F_{-n})_{n \in \mathbb{N}}\) be a decreasing sequence of \(\sigma\)-algebras. Define \(F_{-\infty} := \bigcap_{n \in \mathbb{N}} F_{-n}\). For every \(f \in L^1(B)\),

\[E[f \mid F_{-n}] \xrightarrow{n \to \infty} E[f \mid F_{-\infty}],
\]

almost everywhere and in \(L^1\).

**Proof.** Let

\[V_n = \{g \in L^1(B) \mid E[g \mid F_{-n}] = 0\} ;
\]

for every \(n \in \mathbb{N} \cup \{\infty\}\). Note that from the tower property \(V_1 \subseteq V_2 \subseteq \ldots \subseteq V_\infty\). Define

\[V_\infty = \bigcup_{n \in \mathbb{N}} V_n.
\]

We will prove that \(V_\infty\) is dense in \(V_\infty\). In order to prove it, we use two claims from functional analysis.

1. A corollary of the Hahn-Banach theorem: If for every continuous linear functional \(\Lambda : L^1(B) \to \mathbb{R}\) such that \(V_\infty \subseteq \ker \Lambda\), also \(V_\infty \subseteq \ker \Lambda\), then \(V_\infty\) is dense in \(V_\infty\).
2. Every continuous linear functional $\Lambda : L^1 (B) \to \mathbb{R}$ is of the form

$$\Lambda_h (g) = \int_X g \cdot h d \mu,$$

for $h \in L^\infty (B)$.

Suppose that $V_n \subseteq \ker \Lambda_h$, for every $n \in \mathbb{N}$. Then it follows that

$$\int_X (g - \mathbb{E}[g | \mathcal{F}_{-n}]) h d \mu = 0,$$

for every $g \in L^1 (B)$ and $n \in \mathbb{N}$. In particular, since $L^\infty (B) \subseteq L^1 (B)$, it holds for $g = h$. Thus,

$$\int_X (h - \mathbb{E}[h | \mathcal{F}_{-n}]) h d \mu = 0.$$

Furthermore,

$$\int_X (h - \mathbb{E}[h | \mathcal{F}_{-n}]) \mathbb{E}[h | \mathcal{F}_{-n}] d \mu = \int_X h \mathbb{E}[h | \mathcal{F}_{-n}] d \mu - \int_X \mathbb{E}[h | \mathcal{F}_{-n}] \mathbb{E}[h | \mathcal{F}_{-n}] d \mu = \int_X h \mathbb{E}[h | \mathcal{F}_{-n}] d \mu - \int_X \mathbb{E}[h | \mathcal{F}_{-n}] h d \mu = 0.$$

Subtracting the two equations will give us

$$0 = \int_X (h - \mathbb{E}[h | \mathcal{F}_{-n}]) h d \mu - \int_X (h - \mathbb{E}[h | \mathcal{F}_{-n}]) \mathbb{E}[h | \mathcal{F}_{-n}] d \mu$$

$$= \int_X ((h - \mathbb{E}[h | \mathcal{F}_{-n}]) h - (h - \mathbb{E}[h | \mathcal{F}_{-n}]) \mathbb{E}[h | \mathcal{F}_{-n}]) d \mu$$

$$= \int_X (h^2 - 2h \mathbb{E}[h | \mathcal{F}_{-n}] + (\mathbb{E}[h | \mathcal{F}_{-n}])^2) d \mu$$

$$= \int_X (h - \mathbb{E}[h | \mathcal{F}_{-n}])^2 d \mu.$$

Thus, $h = \mathbb{E}[h | \mathcal{F}_{-n}] \in L^\infty (\mathcal{F}_{-n})$ for every $n \in \mathbb{N}$ and so $h \in L^\infty (\mathcal{F}_{-\infty})$. Finally, let $g \in V_\infty$,

$$\int_X g h d \mu = \int_X \mathbb{E}[g h | \mathcal{F}_{-\infty}] d \mu = \int_X h \mathbb{E}[g | \mathcal{F}_{-\infty}] d \mu = 0.$$

Note that since every $g \in L^1 (B)$ can be written as $g = \mathbb{E}[g | \mathcal{F}_{-\infty}] + (g - \mathbb{E}[g | \mathcal{F}_{-\infty}])$, we conclude that $L^1 (B) = L^1 (\mathcal{F}_{-\infty}) + V_\infty$. Therefore, $L^1 (\mathcal{F}_{-\infty}) + V_\infty$ is dense in $L^1 (B)$.

The theorem clearly holds for functions in $L^1 (\mathcal{F}_{-\infty}) + V_\infty$. The rest of the proof is similar to the proof of Theorem 2.1, and will be left as an exercise (the full proof can be found in pages 131 – 132 in [1]).
4 Corollaries

In this section, we will use the two main theorems that we proved, in order to prove some corollaries in measure theory, ergodic theory and probability theory.

4.1 Measure Theory

**Theorem 4.1 (A Version of Lebesgue Density Theorem).** Let \((X, \mathcal{B}, \mu)\) be a probability space, where \(X = [0, 1)^d\), \(\mathcal{B}\) is the Borel \(\sigma\)-algebra and \(\mu\) is the Lebesgue measure. Define the partition
\[
\xi_n = \left\{ \prod_{i=1}^d \left[ \frac{j_i}{2^n}, \frac{j_i + 1}{2^n} \right] \mid 0 \leq j_i \leq 2^n - 1 \forall 1 \leq i \leq d \right\},
\]
for every \(n \in \mathbb{N}\). Let \(B_n(x)\) be the atom of \(\xi_n\) that contains \(x\), for every \(x \in X\). Let \(A \in \mathcal{B}\), then
\[
\frac{\mu(A \cap B_n(x))}{\mu(B_n(x))} \xrightarrow{n \to \infty} \mathbb{1}_A(x),
\]
for almost every \(x \in X\).

**Proof.** Define \(F_n = \sigma(\xi_n)\) for every \(n \in \mathbb{N}\). Let \(\eta = \mathbb{1}_A\). Note that, \(E[\eta \mid F_n]\) is constant on every \(C \in \xi_n\), and
\[
E[\eta \mid F_n](x) = \frac{\mu(A \cap B_n(x))}{\mu(B_n(x))}.
\]
Thus, by Theorem 2.1,
\[
\frac{\mu(A \cap B_n(x))}{B_n(x)} = E[\eta \mid F_n](x) \xrightarrow{a.e.} E[\eta \mid \mathcal{B}](x) = \eta(x) = 1.
\]

**Remark.** The same proof holds for every increasing sequence of partitions, and probability measure, as long as the union of all the partitions generates the Borel \(\sigma\)-algebra.

4.2 Ergodic Theory

**Theorem 4.2.** Let \(G\) be a LCSC group. Suppose that there’s a sequence of finite subgroups \(G_1 \subseteq G_2 \subseteq \ldots \subseteq \bigcup_{n=1}^\infty G_n =: G_\infty \subseteq G\), such that \(G_\infty\) is dense in \(G\).

Let \((X, \mathcal{B}, \mu)\) be a probability space (\(X\) is compact), and suppose that \(G \curvearrowright X\) continuously, and measure preserving. Let \(G\) be the \(\sigma\)-algebra of \(G_\infty\)-invariant sets which are weakly \(G\)-invariant. Then for every \(f \in L^1(\mathcal{B})\),
\[
f_n(x) := \frac{1}{|G_n|} \sum_{h \in G_n} f(hx) \xrightarrow{n \to \infty} E[f \mid G],
\]
almost everywhere and in \(L^1\).

**Remark.** A good example for a group of this form to keep in mind will be a countable product of finite groups, where \(G_n\) is the product of the first \(n\) groups.

Another example is \(G = (0, 1) \times \mathbb{Z} / (\text{mod } 1)\), with \(G_n = \langle 2^{-n} \rangle\).
Proof. Define $\mathcal{F}_{-n} \subseteq \mathcal{B}$ to be the $\sigma$-algebra of $G_n$-invariant sets. Note that $f_n$ is $\mathcal{F}_{-n}$ measurable. Moreover, for every $A \in \mathcal{F}_{-n}$,
\[
\int_A f_n(x) \, d\mu(x) = \frac{1}{|G_n|} \sum_{h \in G_n} \int_{h^{-1}A} f(y) \, d\mu(y) = \frac{1}{|G_n|} \sum_{h \in G_n} \int_A f(h^{-1}y) \, d\mu(y) = \int_A f(y) \, d\mu(y).
\]
Thus, $f_n = \mathbb{E}[f | \mathcal{F}_{-n}]$. By Theorem 3.2,
\[
f_n(x) \xrightarrow{n \to \infty} \mathbb{E}[f | \mathcal{F}_{-\infty}],
\]
almost everywhere and in $L^1$, where $\mathcal{F}_{-\infty} = \bigcap_{n \in \mathbb{N}} \mathcal{F}_{-n}$. It remains to show that $\mathcal{F}_{-\infty} = \mathcal{G}$.

Let $A \in \mathcal{F}_{-\infty}$ and let $g \in G$. Let $\varepsilon > 0$, fix $f \in C(X)$ such that $\|f - 1_A\|_1 < \varepsilon$. Let $h \in G_\infty$ such that $\|f(gh^{-1}x) - f(x)\|_1 < \varepsilon$, given by the density of $G_\infty$ in $G$, the continuity of the action of $G$, the continuity of $f$ and the dominated convergence theorem. Then,
\[
\mu(A \triangle gA) = \mu \left( \frac{A \triangle gh^{-1}hA}{=A} \right) = \mu(A \triangle gh^{-1}A)
\]
\[
= \int_X \left| \mathbb{1}_A(x) - \mathbb{1}_A(gh^{-1}x) \right| \, d\mu(x)
\]
\[
\leq \int_X \left| \mathbb{1}_A(x) - f(x) \right| \, d\mu(x) + \int_X \left| \mathbb{1}_A(gh^{-1}x) - f(gh^{-1}x) \right| \, d\mu(x)
\]
\[
\leq \int_X \left| f(x) - f(gh^{-1}x) \right| \, d\mu(x) + \int_X \left| f(gh^{-1}y) - f(gh^{-1}x) \right| \, d\mu(y) < 3\varepsilon.
\]
Since $\varepsilon$ is arbitrary, $\mu(A \triangle gA) = 0$, and thus $A$ is weakly $G$-invariant.

Another corollary is the following theorem about conditional measures.

Theorem 4.3. Let $(X, \mathcal{B}, \mu, (\mathcal{F}_n)_{n \in \mathbb{N}})$ be a filtered space (respectively, let $(\mathcal{F}_n)_{n \in \mathbb{N}}$ be a decreasing sequence of $\sigma$-algebra). Suppose $X$ is an LCSC space. Then
\[
\mathbb{E}_{\mu x} \xrightarrow{n \to \infty} \mu_{x \infty} \quad (\text{respectively, } \mathbb{E}_{\mu x}^{-n} \xrightarrow{n \to \infty} \mu_{x \infty}^{-1}),
\]
for $\mu$-almost every $x \in X$.

Proof. We will prove the increasing case. The decreasing case has the same proof, replacing $n$ with $-n$ and $\infty$ with $-\infty$. 

7
Let \((f_k)_k \in \mathbb{N}\) be a dense subset of \(C_c(X)\). In order to prove that \(\mu_F^\infty \xrightarrow{n \to \infty} \mu_F^\infty\), it is sufficient to show that for every \(k \in \mathbb{N}\),
\[
\int_X f_k d\mu_F^\infty - \int_X f_k d\mu_F = 0 \quad \text{as } n \to \infty.
\]
For every \(k \in \mathbb{N}\), let \(E_k\) be a null set, such that for every \(x \in X\setminus E_k\):
1. \(E[f_k | F_n](x) \xrightarrow{n \to \infty} E[f_k | F_\infty](x)\).
2. \(E[f_k | F_n](x) = \int_X f_k d\mu_F^\infty, \forall n \in \mathbb{N} \cup \{\infty\}\).

Define \(X' = X \setminus \left(\bigcup_{k \in \mathbb{N}} E_k\right)\). For every \(x \in X'\) and for every \(k \in \mathbb{N}\),
\[
\int_X f_k d\mu_F^\infty - \int_X f_k d\mu_F = E[f_k | F_n](x) - E[f_k | F_\infty](x) \xrightarrow{n \to \infty} 0,
\]
and \(\mu_F^\infty \xrightarrow{w^*} \mu_F^\infty\).

4.3 Probability Theory

These corollaries won’t be shown in the seminar.

**Theorem 4.4** (Kolmogorov 0–1 Law). Let \((X, \mathcal{B}, \mu)\) be a probability space. Let \((f_n)_{n \in \mathbb{N}}\) be a sequence of independent random variables. Define \(F_n = \sigma(f_1, \ldots, f_n)\) and \(T_n = \sigma(f_{n+1}, f_{n+2}, \ldots)\) for every \(n \geq 0\) (where \(F_0 = \{\emptyset, X\}\)). Let \(T = \bigcap_{n \in \mathbb{N}} T_n\). Then \(\mu(A) \in \{0, 1\}\), for every \(A \in T\).

**Proof.** Let \(A \in T\). Since \(A \in T_0 = F_1\), we know that \(A \in \sigma(F_\infty) = \sigma(\bigcup_{n \in \mathbb{N}} F_n)\).

Define \(\eta = \mathbb{1}_A\). Notice that since \(\eta\) is \(T_n\) measurable, and the random variables are independent, \(\eta\) is independent of \(F_n\), for every \(n \in \mathbb{N}\). Therefore, by Theorem 2.1,
\[
\eta = \mathbb{E}[\eta | F_\infty] = \lim_{n \to \infty} \mathbb{E}[\eta | F_n] = \lim_{n \to \infty} \mathbb{E}[\eta] = \mu(A),
\]
almost surely. The result follows since \(\eta\) only takes values in \(\{0, 1\}\). \(\square\)

**Theorem 4.5** (The Strong Law of Large Numbers). Let \((X, \mathcal{B}, \mu)\) be a probability space. Let \((f_n)_{n \in \mathbb{N}} \subseteq L^1\) be a sequence of IID random variables. Let \(M = \mathbb{E}[f_1]\). Define
\[
S_n = f_1 + \ldots + f_n.
\]
Then \(S_n / n \xrightarrow{n \to \infty} M\) almost everywhere and in \(L^1\).

**Remark.** The usual proof of this theorem gives us only the almost everywhere convergence. This proof will add the convergence in \(L^1\).

**Proof.** Define \(F_n = \sigma(f_n, f_{n+1}, \ldots)\) for every \(n \in \mathbb{N}\), and \(F_\infty = \bigcap_{n \in \mathbb{N}} F_n\). Since \(S_n\) is \(F_n\) measurable, \(\mathbb{E}[S_n | F_n] = S_n\) almost everywhere. By symmetry, \(\mathbb{E}[f_1 | F_\infty] = \frac{S_n}{n}\) almost surely. Hence, by Theorem 3.2,
\[
L = \lim_{n \to \infty} \mathbb{E}[f_1 | F_n] = \lim_{n \to \infty} \frac{S_n}{n},
\]
exists almost everywhere and in $L^1$. Moreover,
\[
\mathbb{E}[L] = \lim_{n \to \infty} \mathbb{E}\left[ \frac{S_n}{n} \right] = M.
\]

Note that
\[
\limsup_{n \to \infty} \frac{f_{k+1} + \ldots + f_{k+n}}{n} = \limsup_{n \to \infty} \frac{f_1 + \ldots + f_{k+n} - f_1 + \ldots + f_k}{n} = \limsup_{n \to \infty} \frac{n+k}{n} \cdot \frac{f_1 + \ldots + f_{k+n}}{n+k} - \frac{f_1 + \ldots + f_k}{n} = L,
\]
almost surely and in $L^1$. Thus, $L$ is $\mathcal{F}_{-k}$ measurable for every $k \in \mathbb{N}$. By Kolmogorov 0–1 law, $L$ is constant almost everywhere, and therefore $L = M$ almost everywhere. 

\textbf{References}