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 $(X, \mathcal{B}, \mu)$ be a measure space, A* filtration *is a sequence of sub-σ-algebras of $\mathcal{B}$, such that $\mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq ... \subseteq \mathcal{B}$. We define:*

$$\mathcal{F}_\infty := \sigma\left(\bigcup_{n\in\mathbb{N}} \mathcal{F}_n\right) \subseteq \mathcal{B}.$$

We will call $\left(X, \mathcal{B}, \mu, (\mathcal{F}_n)_{n\in\mathbb{N}}\right)$ a filtered space*.*

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

1. *$M_n \in L^1(\mathcal{F}_n)$, for every $n \in \mathbb{N}$. I.e., M_n is integrable, and $\mathcal{F}_n$ measurable.*
2. *$\mathbb{E}[M_{n+1} \mid \mathcal{F}_n] = M_n$, 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 $\left(X, \mathcal{B}, \mu, (\mathcal{F}_n)_{n\in\mathbb{N}}\right)$ be a filtered space. Let $f \in L^1(\mathcal{B})$. Define $M_n = \mathbb{E}[f \mid \mathcal{F}_n]$. Note that (M_n) is a Martingale relative to $(\mathcal{F}_n)$. Indeed:*

- *By the definition of the conditional expectation, the process M_n is adapted to $(\mathcal{F}_n)$, and $M_n \in L^1$.*
- *By the tower property,*

$$\mathbb{E}[M_{n+1} \mid \mathcal{F}_n] = \mathbb{E}[\mathbb{E}[f \mid \mathcal{F}_{n+1}] \mid \mathcal{F}_n] = [f \mid \mathcal{F}_n] = M_n.$$

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:

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

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 μ be the Lebesgue measure. Define the filtration*

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

and the process $M_n = (n+1)\,\mathbb{1}_{[0,\frac{1}{n+1})}$, 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 $M_n \xrightarrow[n\to\infty]{a.s.} 0$. However,

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

and therefore M_n does not converge to M_∞ 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 $\left(X, \mathcal{B}, \mu, (\mathcal{F}_n)_{n\in\mathbb{N}}\right)$ 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*

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

Then

$$\mu(E) \le \frac{1}{\lambda}\|f\|_1.$$

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

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

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

$$E_n^N = \left\{ x \in X \mid \mathbb{E}\left[f \mid \mathcal{F}_n\right](x) > \lambda \text{ and } \mathbb{E}\left[f \mid \mathcal{F}_i\right](x) \leq \lambda \text{ for } 1 \leq i \leq n-1 \right\}.$$

Note that $E^N = \biguplus_{n=1}^{N} E_n^N$, and $E_n^N \in \mathcal{F}_n$ since $\mathcal{F}_1, \mathcal{F}_2, ..., \mathcal{F}_{n-1} \subseteq \mathcal{F}_n$. Thus,

$$\begin{aligned} \lambda\mu\left(E^N\right) = \sum_{n=1}^{N} \lambda\mu\left(E_n^N\right) &= \sum_{n=1}^{N} \int_{E_n^N} \lambda \\ &\leq \sum_{n=1}^{N} \int_{E_n^N} \mathbb{E}\left[f \mid \mathcal{F}_n\right] \\ &= \sum_{n=1}^{N} \int_{E_n^N} f = \int_{E^N} f \leq \|f\|_1. \end{aligned}$$

By measure continuity, we deduce

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

$\square$

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

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

$$\left\{ A \in \mathcal{B} \mid \forall \varepsilon > 0 \, \exists \left(m \in \mathbb{N} \wedge A_m \in \mathcal{F}_m\right) \text{ such that } \mu\left(A \triangle A_m\right) < \varepsilon \right\},$$

is a σ-algebra that contains $\bigcup_{n \in \mathbb{N}} \mathcal{F}_n$ and thus contains $\mathcal{F}_\infty$.

Proof of Theorem 2.1. Since $\mathbb{E}\left[\mathbb{E}\left[f \mid \mathcal{F}_\infty\right] \mid \mathcal{F}_n\right] = \mathbb{E}\left[f \mid \mathcal{F}_n\right]$, we may assume without loss of generality that f is $\mathcal{F}_\infty$ measurable.
If there exists $m \in \mathbb{N}$ such that f is $\mathcal{F}_m$ measurable, then $\left[f \mid \mathcal{F}_n\right]$ is eventually constant sequence and thus the theorem holds. From Lemma 2.3, for every $\varepsilon > 0$, there exists $m \in \mathbb{N}$ and $g \in L^1\left(\mathcal{F}_m\right)$ such that $\|f - g\|_1 < \frac{\varepsilon}{2}$. Thus, for every $n \geq m$,

$$\left\|\mathbb{E}\left[f \mid \mathcal{F}_n\right] - f\right\|_1 \leq \left\|\mathbb{E}\left[f \mid \mathcal{F}_n\right] - \mathbb{E}\left[g \mid \mathcal{F}_n\right]\right\|_1 + \|f - g\|_1 + \underbrace{\left\|\mathbb{E}\left[g \mid \mathcal{F}_n\right] - g\right\|_1}_{=0} < \varepsilon.$$

For the almost everywhere convergence, we note that

$$\begin{aligned}
&\mu\left(\left\{x\in X \mid \limsup_{n\to\infty}\left|\mathbb{E}\left[f\mid\mathcal{F}_n\right]-f\right|(x)>\sqrt{\varepsilon}\right\}\right)\\
&=\mu\left(\left\{x\in X \mid \limsup_{n\to\infty}\left|\mathbb{E}\left[f-g\mid\mathcal{F}_n\right](x)-(f-g)(x)\right|>\sqrt{\varepsilon}\right\}\right)\\
&\leq\mu\left(\left\{x\in X \mid \sup_{n\in\mathbb{N}}\mathbb{E}\left[|f-g|\mid\mathcal{F}_n\right](x)>\frac{\sqrt{\varepsilon}}{2}\right\}\right)+\mu\left(\left\{x\in X \mid |f-g|(x)>\frac{\sqrt{\varepsilon}}{2}\right\}\right)\\
&\overset{Lemma\ 2.2}{\leq}\frac{2}{\sqrt{\varepsilon}}\|f-g\|_1+\frac{2}{\sqrt{\varepsilon}}\|f-g\|_1\leq 2\sqrt{\varepsilon}.
\end{aligned}$$

Since ε is arbitrary, it follows that $\limsup_{n\to\infty}\left|\mathbb{E}\left[f\mid\mathcal{F}_n\right]-f\right|=0$ almost everywhere. ☐

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 $(\mathcal{F}_{-n})_{n\in\mathbb{N}}$ be a decreasing sequence of σ-algebras. I.e.,*

$$\mathcal{B}\supseteq\mathcal{F}_{-1}\supseteq\mathcal{F}_{-2}\supseteq\ldots\supseteq\bigcap_{n\in\mathbb{N}}\mathcal{F}_{-n}=:\mathcal{F}_{-\infty}.$$

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

1. *$M_n\in L^1(\mathcal{F}_{-n})$, for every $n\in\mathbb{N}$.*
2. *$\mathbb{E}\left[M_{-n+1}\mid\mathcal{F}_{-n}\right]=M_{-n}$, for every $n\in\mathbb{N}$.*

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

$$\mathbb{E}\left[f\mid\mathcal{F}_{-n}\right]\xrightarrow[n\to\infty]{}\mathbb{E}\left[f\mid\mathcal{F}_{-\infty}\right],$$

almost everywhere and in L^1.

Proof. Let

$$V_n=\left\{g\in L^1(\mathcal{B})\mid\mathbb{E}\left[g\mid\mathcal{F}_{-n}\right]=0\right\},$$

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

$$V_*=\bigcup_{n\in\mathbb{N}}V_n.$$

We will prove that V_* is dense in V_∞. 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 $\Lambda: L^1(\mathcal{B})\to\mathbb{R}$ such that $V_*\subseteq\ker\Lambda$, also $V_\infty\subseteq\ker\Lambda$, then V_* is dense in V_∞.

2. Every continuous linear functional $\Lambda : L^1(\mathcal{B}) \to \mathbb{R}$ is of the form

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

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

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

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

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

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

Furthermore,

$$\begin{aligned}
&\int_X (h - \mathbb{E}[h \mid \mathcal{F}_{-n}]) \mathbb{E}[h \mid \mathcal{F}_{-n}] d\mu \\
&= \int_X h\mathbb{E}[h \mid \mathcal{F}_{-n}] d\mu - \int_X \mathbb{E}[h \mid \mathcal{F}_{-n}] \mathbb{E}[h \mid \mathcal{F}_{-n}] d\mu \\
&= \int_X h\mathbb{E}[h \mid \mathcal{F}_{-n}] d\mu - \int_X \mathbb{E}[\mathbb{E}[h \mid \mathcal{F}_{-n}] h \mid \mathcal{F}_{-n}] d\mu \\
&= \int_X h\mathbb{E}[h \mid \mathcal{F}_{-n}] d\mu - \int_X \mathbb{E}[h \mid \mathcal{F}_{-n}] h d\mu = 0.
\end{aligned}$$

Subtracting the two equations will give us

$$\begin{aligned}
0 &= \int_X (h - \mathbb{E}[h \mid \mathcal{F}_{-n}]) h d\mu - \int_X (h - \mathbb{E}[h \mid \mathcal{F}_{-n}]) \mathbb{E}[h \mid \mathcal{F}_{-n}] d\mu \\
&= \int_X ((h - \mathbb{E}[h \mid \mathcal{F}_{-n}]) h - (h - \mathbb{E}[h \mid \mathcal{F}_{-n}]) \mathbb{E}[h \mid \mathcal{F}_{-n}]) d\mu \\
&= \int_X \left(h^2 - 2h\mathbb{E}[h \mid \mathcal{F}_{-n}] + (\mathbb{E}[h \mid \mathcal{F}_{-n}])^2\right) d\mu \\
&= \int_X (h - \mathbb{E}[h \mid \mathcal{F}_{-n}])^2 d\mu.
\end{aligned}$$

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

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

Note that since every $g \in L^1(\mathcal{B})$ can be written as $g = \mathbb{E}[g \mid \mathcal{F}_{-\infty}] + (g - \mathbb{E}[g \mid \mathcal{F}_{-\infty}])$, we conclude that $L^1(\mathcal{B}) = L^1(\mathcal{F}_{-\infty}) + V_\infty$. Therefore, $L^1(\mathcal{F}_{-\infty}) + V_*$ is dense in $L^1(\mathcal{B})$.
The theorem clearly holds for functions in $L^1(\mathcal{F}_{-\infty}) + V_*$. 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]). $\square$

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 σ-algebra and μ is the Lebesgue measure. Define the partition*

$$\xi_n = \left\{ \prod_{i=1}^{d} \left[\frac{j_i}{2^n}, \frac{j_i+1}{2^n} \right) \mid 0 \le j_i \le 2^n - 1 \forall 1 \le i \le d \right\},$$

for every $n \in \mathbb{N}$. Let $B_n(x)$ be the atom of ξ_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 $\mathcal{F}_n = \sigma(\xi_n)$ for every $n \in \mathbb{N}$. Let $\eta = \mathbb{1}_A$. Note that, $\mathbb{E}[\eta \mid \mathcal{F}_n]$ is constant on every $C \in \xi_n$, and

$$\mathbb{E}[\eta \mid \mathcal{F}_n](x) = \frac{\mu(A \cap B_n(x))}{\mu(B_n(x))}.$$

Thus, by Theorem 2.1,

$$\frac{\mu(A \cap B_n(x))}{B_n(x)} = \mathbb{E}[\eta \mid \mathcal{F}_n](x) \xrightarrow[n\to\infty]{a.e.} \mathbb{E}[\eta \mid \mathcal{B}](x) = \eta(x) = 1.$$

□

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 σ-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_∞ 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 $\mathcal{G}$ be the σ-algebra of G_∞-invariant sets which are weakly G-invariant. Then for every $f \in L^1(\mathcal{B})$,

$$f_n(x) := \frac{1}{|G_n|} \sum_{h \in G_n} f(hx) \xrightarrow[n\to\infty]{} \mathbb{E}[f \mid \mathcal{G}],$$

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 $G = \left([0,1), +_{\text{(mod 1)}}\right)$, with $G_n = \langle 2^{-n} \rangle$.

Proof. Define $\mathcal{F}_{-n} \subseteq \mathcal{B}$ to be the σ-algebra of G_n-invariant sets. Note that f_n is $\mathcal{F}_{-n}$ measurable. Moreover, for every $A \in \mathcal{F}_{-n}$,

$$\begin{aligned}\int_A f_n(x)\,d\mu(x) &= \frac{1}{|G_n|}\sum_{h\in G_n}\int_A f(hx)\,d\mu(x) = [y = hx] \\ &= \frac{1}{|G_n|}\sum_{h\in G_n}\int_{h^{-1}A} f(y)\,d\mu(y) = \\ &= \frac{1}{|G_n|}\sum_{h\in G_n}\int_A f(y)\,d\mu(y) = \int_A f(y)\,d\mu(y).\end{aligned}$$

Thus, $f_n = \mathbb{E}[f \mid \mathcal{F}_{-n}]$. By Theorem 3.2,

$$f_n(x) \xrightarrow[n\to\infty]{} \mathbb{E}[f \mid \mathcal{F}_{-\infty}],$$

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 $\|f - \mathbb{1}_A\|_1 < \varepsilon$. Let $h \in G_\infty$ such that $\left\|f(gh^{-1}x) - f(x)\right\|_1 < \varepsilon$, given by the density of G_∞ in G, the continuity of the action of G, the continuity of f and the dominated convergence theorem. Then,

$$\begin{aligned}\mu(A\triangle gA) &= \mu\left(A\triangle gh^{-1}\underbrace{hA}_{=A}\right) = \mu(A\triangle gh^{-1}A) \\ &= \int_X \left|\mathbb{1}_A(x) - \mathbb{1}_A(hg^{-1}x)\right| d\mu(x) \\ &\le \underbrace{\int_X |\mathbb{1}_A(x) - f(x)|\,d\mu(x)}_{<\varepsilon} + \underbrace{\int_X |\mathbb{1}_A(hg^{-1}x) - f(hg^{-1}x)|\,d\mu(x)}_{<\varepsilon} \\ &+ \int_X \left|f(x) - f(hg^{-1}x)\right| d\mu(x) \\ &\overset{[y=hg^{-1}x]}{<} 2\varepsilon + \underbrace{\int_X |f(gh^{-1}y) - f(y)|\,d\mu(y)}_{<\varepsilon} < 3\varepsilon.\end{aligned}$$

Since ε 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 $\left(X, \mathcal{B}, \mu, (\mathcal{F}_n)_{n\in\mathbb{N}}\right)$ be a filtered space (respectively, let $(\mathcal{F}_{-n})_{n\in\mathbb{N}}$ be a decreasing sequence of σ-algebra). Suppose X is an LCSC space. Then*

$$\mu_x^{\mathcal{F}_n} \xrightarrow[n\to\infty]{w^*} \mu_x^{\mathcal{F}_\infty} \text{ (respectively, } \mu_x^{\mathcal{F}_{-n}} \xrightarrow[n\to\infty]{w^*} \mu_x^{\mathcal{F}_{-\infty}}\text{)},$$

for μ-almost every $x \in X$.

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

Let $(f_k)_k \in \mathbb{N}$ be a dense subset of $C_c(X)$. In order to prove that $\mu_x^{\mathcal{F}_n} \xrightarrow[n\to\infty]{w^*} \mu_x^{\mathcal{F}_\infty}$, it is sufficient to show that for every $k \in \mathbb{N}$,

$$\int_X f_k d\mu_x^{\mathcal{F}_n} - \int_X f_k d\mu_x^{\mathcal{F}_\infty} \xrightarrow[n\to\infty]{} 0.$$

For every $k \in \mathbb{N}$, let E_k be a null set, such that for every $x \in X \backslash E_k$:

1. $\mathbb{E}[f_k \mid \mathcal{F}_n](x) \xrightarrow[n\to\infty]{} \mathbb{E}[f_k \mid \mathcal{F}_\infty](x)$.
2. $\mathbb{E}[f_k \mid \mathcal{F}_n](x) = \int_X f_k d\mu_x^{\mathcal{F}_n}$, for every $n \in \mathbb{N} \cup \{\infty\}$.

Define $X' = X \backslash \left(\bigcup_{k\in\mathbb{N}} E_k\right)$. For every $x \in X'$ and for every $k \in \mathbb{N}$,

$$\int_X f_k d\mu_x^{\mathcal{F}_n} - \int_X f_k d\mu_x^{\mathcal{F}_\infty} = \mathbb{E}[f_k \mid \mathcal{F}_n](x) - \mathbb{E}[f_k \mid \mathcal{F}_\infty](x) \xrightarrow[n\to\infty]{} 0,$$

and $\mu_x^{\mathcal{F}_n} \xrightarrow[n\to\infty]{w^*} \mu_x^{\mathcal{F}_\infty}$. □

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 $\mathcal{F}_n = \sigma(f_1, ..., f_n)$ and $\mathcal{T}_n = \sigma(f_{n+1}, f_{n+2}, ...)$ for every $n \geq 0$ (where $\mathcal{F}_0 = \{\emptyset, X\}$). Let $\mathcal{T} = \bigcap_{n\in\mathbb{N}} \mathcal{T}_n$. Then $\mu(A) \in \{0, 1\}$, for every $A \in \mathcal{T}$.*

Proof. Let $A \in \mathcal{T}$. Since $A \in \mathcal{T}_0 = \mathcal{F}_1$, we know that $A \in \sigma(\mathcal{F}_\infty) = \sigma\left(\bigcup_{n\in\mathbb{N}} \mathcal{F}_n\right)$.
Define $\eta = \mathbb{1}_A$. Notice that since η is $\mathcal{T}_n$ measurable, and the random variables are independent, η is independent of $\mathcal{F}_n$, for every $n \in \mathbb{N}$. Therefore, by Theorem 2.1,

$$\eta = \mathbb{E}[\eta \mid \mathcal{F}_\infty] = \lim_{n\to\infty} \mathbb{E}[\eta \mid \mathcal{F}_n] = \lim_{n\to\infty} \mathbb{E}[\eta] = \mu(A),$$

almost surely. The result follows since η only takes values in $\{0, 1\}$. □

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*

$$S_n = f_1 + ... + f_n.$$

Then $\frac{S_n}{n} \xrightarrow[n\to\infty]{} M$ 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 $\mathcal{F}_{-n} = \sigma(f_n, f_{n+1}, ...)$ for every $n \in \mathbb{N}$, and $\mathcal{F}_{-\infty} = \bigcap_{n\in\mathbb{N}} \mathcal{F}_{-n}$. Since S_n is $\mathcal{F}_{-n}$ measurable, $\mathbb{E}[S_n \mid \mathcal{F}_{-n}] = S_n$ almost everywhere. By symmetry, $\mathbb{E}[f_1 \mid \mathcal{F}_{-n}] = \frac{S_n}{n}$ almost surely.
Hence, by Theorem 3.2,

$$L = \lim_{n\to\infty} \mathbb{E}[f_1 \mid \mathcal{F}_{-n}] = \lim_{n\to\infty} \frac{S_n}{n},$$

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

$$\begin{aligned}\limsup_{n\to\infty} \frac{f_{k+1} + \dots + f_{k+n}}{n} &= \limsup_{n\to\infty} \frac{f_1 + \dots + f_{k+n}}{n} - \frac{f_1 + \dots + f_k}{n} \\ &= \limsup_{n\to\infty} \frac{n+k}{n} \cdot \frac{f_1 + \dots + f_{k+n}}{n+k} - \frac{f_1 + \dots + f_k}{n} = L,\end{aligned}$$

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. □

References

[1] Einsiedler M. and Ward T. (2011), Ergodic Theory: with a view towards Number Theory, Springer, London.