LECTURE 6

RUNLIN ZHANG†

Contents

Notation

1. Lecture 6, conditional measures

1.1. **Prelude.** In probability theory, one often has a space, thought of as the collection of all possible "events" together with a probability measure, measuring which event is more likely to happen. Given these data, one can make predictions on "random variables". In mathematical terms, a random variable is just a (measurable) function and the "expectation" of this function is nothing but its integration.

Conditional expectations of a random variable means that we make predictions based on certain information. For instance, one might have another function g on this space and we have perfect knowledge of what the value of g is. So "conditional on" the value of g taken, we make more refined predictions on our random variable.

Conditional expectations, just like expectations, can also be written as integration of the random variable against certain probability measures, known as conditional measures.

From a different perspective, one may also view conditional measures as "Fubini-type theorem".

The material of this lecture is mostly taken from [EW11, chapter 5].

1.2. **Statement of the main theorem.** Let X be a compact metrizable topological space and $\mathscr{C}_X$ be its Borel σ-algebra. Let μ be a probability measure on $(X, \mathscr{B}_X)$. We refer the triple $(X, \mathscr{B}_X, \mu)$ as a compact Borel probability space.

Theorem 1.1. *Let $(X, \mathscr{B}_X, \mu)$ be a compact Borel probability space and $\mathscr{A} \subset \mathscr{B}_X$ be a σ-subalgebra. Then there exist a subset $X' \in \mathscr{A}$ of full μ-measure (i.e. $\mu(X \setminus X') = 0$) and a map $X' \to \mathrm{Prob}(X, \mathscr{B}_X)$, denoted by $x \mapsto \mu_x^{\mathscr{A}}$, satisfying:*

(1) *for every $f \in C(X)$, the map $x \mapsto \int_X f(\omega)\, \mu_x^{\mathscr{A}}(\omega)$ from X' to $\mathbb{R}$ is measurable (w.r.t. $\mathscr{A} \cap X'$) and*

$$\int_{A \cap X'} \left(\int_X f(\omega)\, \mu_x^{\mathscr{A}}(\omega) \right) \mu(x) = \int_A f(\omega)\, \mu(\omega), \quad \forall\, A \in \mathscr{A}.$$

† Email: zhangrunlinmath@outlook.com.

(2) *for every $E \in \mathscr{B}_X$,*

$$x \mapsto \int_X \mathbf{1}_E(\omega)\, \mu_x^{\mathscr{A}}(\omega)$$

is measurable on $(X', \mathscr{A} \cap X')$ and

$$\int_{A\cap X'} \left(\int_X \mathbf{1}_E(\omega)\, \mu_x^{\mathscr{A}}(\omega) \right) \mu(x) = \int_A \mathbf{1}_E(\omega)\, \mu(\omega), \quad \forall A \in \mathscr{A}.$$

or in different terms,

$$\int_{A\cap X'} \mu_x^{\mathscr{A}}(E)\, \mu(x) = \mu(A \cap E), \quad \forall A \in \mathscr{A}.$$

(3) *If $Y \in \mathscr{A}$ is of full measure and $x \mapsto \nu_x^{\mathscr{A}}$ is another map from Y to $\mathrm{Prob}(X, \mathcal{B}_X)$ satisfying for every f in some dense subset of $C(X)$, the map $x \mapsto \int_X f(\omega)\, \mu_x^{\mathscr{A}}(\omega)$ from Y to $\mathbb{R}$ is measurable (w.r.t. $\mathscr{B}_X \cap Y$) and*

$$\int_{A\cap Y} \left(\int_X f(\omega)\, \nu_x^{\mathscr{A}}(\omega) \right) \mu(x) = \int_A f(\omega)\, \mu(\omega), \quad \forall A \in \mathscr{A},$$

then there exists $Y' \subset X' \cap Y$ in $\mathscr{A}$ of full measure such that $\mu_x^{\mathscr{A}} = \nu_x^{\mathscr{A}}$ for all $x \in Y'$;

(4) *If $\mathscr{A}$ is additionally assumed to be countably generated, then one can choose $X'' \subset X'$ in $\mathscr{A}$ of full measure such that $\mu_x^{\mathscr{A}}([x]_{\mathscr{A}}) = 1$ [1] for every $x \in X''$ and $\mu_y^{\mathscr{A}} = \mu_x^{\mathscr{A}}$ whenever $[x]_{\mathscr{A}} = [y]_{\mathscr{A}} \subset X''$.*

(5) *If $\mathscr{A}_1 \subset \mathscr{A}_2 \subset ...$ is an increasing sequence of σ-subalgebras and $\mathscr{A}_\infty$ is the smallest σ-subalgebra containing all of them, then for every $E \in \mathscr{B}$, for μ-almost all x, the relevant conditional measures are defined and*

$$\lim_{n\to\infty} \mu_x^{\mathscr{A}_n}(E) = \mu_x^{\mathscr{A}_\infty}(E).$$

The family of measures $(\mu_x^{\mathscr{A}})$ satisfying condition (1) and (2) as in the theorem are referred to as **conditional measures**.

There are two examples when the conclusion of the theorem (which we leave to the reader to fill in) might be more familiar to the reader.

Example 1.2. *$\mathscr{A}$ is generated by a finite measurable partition $(P_1, P_2, ..., P_n)$ of X. If you prefer, you may even take X to be a finite set to see what happens.*

Example 1.3. *$X = [0,1]^2$ and $\mu = \phi(x,y)\mathrm{dxdy}$ where $\phi(x,y)$ is a measurable non-negative function with $\int \phi(x,y)\mathrm{dxdy} = 1$. And*

$$\mathscr{A} := \{A \times [0,1],\ A \text{ is Borel measurable in } [0,1]\}.$$

1.3. **The set X'.** We are going to construct the measure, thanks to Riesz representation theorem, by specifying the integrals of continuous functions.

First, we choose a countable dense subset $\mathscr{C} \subset C(X)$ containing the constant one function. Let

$$\mathscr{C}_{\mathbb{Q}} := \{ \text{ finite } \mathbb{Q}\text{-linear combinations of elements in } \mathscr{C} \}$$

Thus $\mathscr{C}_{\mathbb{Q}}$ is a countable dense $\mathbb{Q}$-linear subspace of $C(X)$. Let $\pi_{\mathscr{A}}$ denote the orthogonal projection from $L^2(X, \mathscr{B}_X, \mu) \to L^2(X, \mathscr{A}, \mu)$. For every $f \in \mathscr{C}_{\mathbb{Q}}$, choose some representative $f^{\mathscr{A}}$ of $\pi_{\mathscr{A}}([f])$ [2]. Without loss of generality, for the constant one function $\mathbf{1}_X$, which is $\mathscr{A}$-measurable, choose $\mathbf{1}_X^{\mathscr{A}} := \mathbf{1}_X$.

Lemma 1.4. *For every $f \in \mathscr{C}_{\mathbb{Q}}$,*

$$\mu \left\{ x \,\middle|\, \left| f^{\mathscr{A}}(x) \right| > \|f\|_{\sup} \right\} = 0.$$

Proof. The sets $A^\star := \left\{ x \,\middle|\, \star f^{\mathscr{A}}(x) > \|f\|_{\sup} \right\}$ for $\star = +$ or $-$ are in $\mathscr{A}$. Thus their characteristic functions $\mathbf{1}_{A^\star}$ are contained in $L^2(X, \mathscr{B}_X, \mu)$. So if $\mu(A^+) \neq 0$,

$$\|f\|_{\sup} \mu(A^+) \geq \langle [f], \mathbf{1}_{A^+} \rangle = \langle \pi_{\mathscr{A}}([f]), \mathbf{1}_{A^+} \rangle > \|f\|_{\sup} \mu(A^+),$$

a contradiction. So $\mu(A^+) = 0$. Similarly, $\mu(A^-) = 0$ and hence $\mu(A) = 0$. □

[1] See Section 1.7 for the definition of $[x]_{\mathscr{A}}$.

[2] In order to distinguish a genuine function f from its equivalence class up to measure zero, we write $[f]$ for the equivalence class.

Similarly, one shows that

Lemma 1.5. *For every $f \in \mathscr{C}_{\mathbb{Q}}$ with $f \geq 0$, one has $\{x,\ f^{\mathscr{A}}(x) \geq 0\}$ is an element of $\mathscr{A}$ with full measure.*

Lemma 1.6. *For every finite collection $(f_0, f_1, ..., f_n) \subset \mathscr{C}_{\mathbb{Q}}$ and finitely many $(q_1, ..., q_n) \subset \mathbb{Q}$ such that*

$$f_0 = \sum q_i f_i,$$

the set

$$\left\{ x \;\middle|\; f_0^{\mathscr{A}}(x) = \sum q_i f_i^{\mathscr{A}}(x) \right\}$$

is $\mathscr{A}$-measurable and has full measure.

As there are only countably many data, we can find a $\mathscr{A}$-measurable set X' of full measure such that for every $x \in X'$,

(0) $\mathbf{1}_X^{\mathscr{A}}(x) = 1$;
(1) $\left|f^{\mathscr{A}}(x)\right| \leq \|f\|_{\sup}, \quad \forall\, f \in \mathscr{C}_{\mathbb{Q}}$;
(2) $f^{\mathscr{A}}(x) \geq 0, \quad \forall\, f \in \mathscr{C}_{\mathbb{Q}},\ f \geq 0$;
(3) $f_0^{\mathscr{A}}(x) = \sum_{i=1}^n q_i f_i^{\mathscr{A}}(x), \quad \forall\, (f_i)_{i=0}^n \subset \mathscr{C}_{\mathbb{Q}},\ (q_i) \subset \mathbb{Q}$ with $f_0 = \sum_{i=1}^n q_i f_i$.

1.4. **Construction of measures.** For every $x \in X'$ and $f \in C(X)$, find $(f_n) \subset \mathscr{C}_{\mathbb{Q}}$ converging to f in sup-norm. We define $\Lambda_x : C(X) \to \mathbb{R}$ by $\Lambda_x(f) := \lim_{n\to\infty} f_n^{\mathscr{A}}(x)$.

Lemma 1.7. *For $x \in X'$, $(f_n^{\mathscr{A}}(x))$ converges. Consequently, $\Lambda_x(f)$ is well-defined and independent of the choice of (f_n).*

Proof. Take $n, m \in \mathbb{Z}^+$ with $\|f_n - f_m\|_{\sup} \leq \varepsilon$. As $f_n - f_m \in \mathscr{C}$, we have

$$\left|f_n^{\mathscr{A}}(x) - f_m^{\mathscr{A}}(x)\right| = \left|(f_n - f_m)^{\mathscr{A}}(x)\right| \leq \|f_n - f_m\|_{\sup} \leq \varepsilon.$$

This shows that $(f_n^{\mathscr{A}}(x))$ is a Cauchy sequence. □

Also, one sees from the lemma that $\Lambda_x(f)$ is independent of the choice of (f_n). Moreover,

Lemma 1.8. *For $x \in X'$, Λ_x defines a positive bounded linear functional on $C(X)$ sending $\mathbf{1}$ to 1. Therefore, by Riesz representation theorem, there exists a unique Borel probability measure, denoted as $\mu_x^{\mathscr{A}}$, such that $\Lambda_x(f) = \int f(\omega)\mu_x^{\mathscr{A}}(\omega)$.*

Part (1) of Theorem 1.1 is automatically true for $f \in \mathscr{C}_{\mathbb{Q}}$, the general case follows by, say, dominated convergence theorem.

1.5. **Extending to measurable functions.** Let O be an open subset of X, by Urysohn lemma (see e.g. 2.12 of [Rud87, Chapter 2]), there exists a sequence of continuous functions (f_n) that is uniformly bounded and converges to $\mathbf{1}_O$. Similarly, one can find a uniformly bounded sequence of continuous functions converging to the characteristic function of a closed subset.

By dominated convergence theorem, Part (2) of Theorem 1.1 holds for E being open or compact. Actually, the characteristic function of $E = O \cap C$, the intersection of some open subset and closed subset (for simplicity, we shall call such a set **locally closed**), can also be pointwisely approximated by a sequence of uniformly bounded continuous functions. Let

$$\mathcal{R} := \{\text{subsets that can be written as a finite disjoint union of locally closed subsets}\}.$$

Lemma 1.9. *$\mathcal{R}$ is an algebra in the sense that it is closed under taking complements, finite intersections and finite unions.*

Proof. For C_1, C_2 closed and O_1, O_2 open, we note that $(C_1 \cap O_1) \cup (C_2 \cap O_2)$ is a disjoint union of locally closed subsets:

$$\begin{aligned}
&(C_1 \cap O_1) \cup (C_2 \cap O_2) \\
=\,& ((C_1 \cap O_1) \cap (C_2 \cap O_2)) \sqcup (C_1 \cap O_1) \cap (C_2 \cap O_2)^c \\
=\,& ((C_1 \cap C_2) \cap (O_1 \cap O_2)) \sqcup ((C_1 \cap O_1) \cap (C_2^c \cup O_2^c)) \\
=\,& ((C_1 \cap C_2) \cap (O_1 \cap O_2)) \sqcup (C_1 \cap O_1 \cap C_2^c) \sqcup (C_1 \cap O_1 \cap O_2^c \cap C_2).
\end{aligned}$$

The rest follows from this. □

On the other hand, let

$$\mathcal{M} := \{\text{subsets of } \mathscr{B}_X \text{ that satisfy part (2) of Theorem 1.1}\}$$

Lemma 1.10. *Let $E_1 \subset E_2 \subset \ldots$ be an increasing sequence of elements in $\mathcal{M}$, then $E_\infty := \bigcup E_i$ belongs to $\mathcal{M}$. If $E \in \mathcal{M}$, then $E^c \in \mathcal{M}$.*

Proof. This follows from dominated convergence theorem. □

Let $\sigma(\mathcal{R})$ denote the smallest σ-subalgebra of $\mathcal{B}_X$ containing $\mathcal{R}$. We have shown that $\mathcal{R} \subset \mathcal{M}$. It is a general fact that if $\mathcal{M}$ is a subset of some σ-algebra satisfying the conclusion of Lemma 1.10 and contains some subalgebra $\mathcal{R}$ as in Lemma 1.9, then $\mathcal{M}$ contains $\sigma(\mathcal{R})$. In the current case, $\sigma(\mathcal{R})$ is $\mathcal{B}_X$, so they are equal.

Lemma 1.11. *$\mathcal{M} = \sigma(\mathcal{R})$.*

Proof. Let $\mathcal{M}_0$ be the smallest subset of $\mathcal{M}$ containing $\mathcal{R}$ such that the conclusion of Lemma 1.10 holds.

First take $E \in \mathcal{R} \subset \mathcal{M}_0$, consider

$$\mathcal{M}_E := \{F \in \mathcal{M}_0 \mid E \cap F,\, E \cup F,\, E^c \cap F,\, E^c \cup F \in \mathcal{M}_0\}$$

If $F_1 \subset F_2 \subset \ldots$ are contained in $\mathcal{M}_E$, then

$$\begin{aligned} E \cap (\bigcup F_i) &= \bigcup (E \cap F_i), \quad E \cup (\bigcup F_i) = \bigcup (E \cup F_i), \\ E^c \cap (\bigcup F_i) &= \bigcup (E^c \cap F_i), \quad E^c \cup (\bigcup F_i) = \bigcup (E^c \cup F_i) \end{aligned}$$

are all contained in $\mathcal{M}_0$. Hence $\cup F_i \in \mathcal{M}_E$. If $F \in \mathcal{M}_E$, then the complements of

$$E \cap F^c,\, E \cup F^c,\, E^c \cap F^c,\, E^c \cup F^c$$

are contained $\mathcal{M}_0$. Thus they are also contained in $\mathcal{M}_0$, implying that $F^c \in \mathcal{M}_E$.

So we have shown that $\mathcal{M}_E$ satisfies the conclusion of Lemma 1.10. On the other hand, $\mathcal{M}_E$ contains $\mathcal{R}$ by Lemma 1.9. By minimality of $\mathcal{M}_0$, we get $\mathcal{M}_E = \mathcal{M}_0$.

For general $E \in \mathcal{M}_0$, $\mathcal{M}_F = \mathcal{M}_0$, $\forall F \in \mathcal{R}$ implies that $\mathcal{R} \subset \mathcal{M}_E$. Same arguments as above show that $\mathcal{M}_E$ satisfies the conclusion of Lemma 1.10. Again by minimality, $\mathcal{M}_E = \mathcal{M}_0$.

Now that $\mathcal{M}_0$ is closed under taking finite unions and intersections, one can directly verify that $\mathcal{M}_0$ is a σ-algebra. This forces $\mathcal{M}_0 = \mathcal{M} = \sigma(\mathcal{R})$. □

Remark 1.12. *Similar arguments are used to prove the π-λ theorem in measure theory.*

1.6. **Uniqueness.** Now we turn to part (3) of Theorem 1.1.

So we have a dense subset $\mathscr{C}$ of $C(X)$ such that for every $f \in \mathscr{C}$ and $A \in \mathscr{A}$

$$\int_{A\cap Y} \left(\int_X f(\omega)\, \nu_x^{\mathscr{A}}(\omega) \right) \mu(x) = \int_A f(\omega)\, \mu(\omega) = \int_{A\cap Y} \left(\int_X f(\omega)\, \mu_x^{\mathscr{A}}(\omega) \right) \mu(x). \tag{1}$$

Let $\mathscr{C}' \subset \mathscr{C}$ be a countable subset that is still dense in $C(X)$. For each $f \in \mathscr{C}'$, let

$$\begin{aligned} D_f^+ &:= \left\{x \in X' \cap Y \;\middle|\; \mu_x^{\mathscr{A}}(f) > \nu_x^{\mathscr{A}}(f)\right\} \\ D_f^- &:= \left\{x \in X' \cap Y \;\middle|\; \mu_x^{\mathscr{A}}(f) < \nu_x^{\mathscr{A}}(f)\right\}. \end{aligned}$$

Applying Equa.(1) we see that both D_f^+ and D_f^- has measure zero. Let Y' be the complement of their unions as f varies in $\mathscr{C}'$. Then Y' is full in X. And for every $x \in Y'$ and every $f \in \mathscr{C}'$,

$$\int_X f(\omega)\, \mu_x^{\mathscr{A}}(\omega) = \int_X f(\omega)\, \nu_x^{\mathscr{A}}(\omega)$$

which extends to all $f \in \mathscr{C}'$ by dominated convergence theorem. So $\mu_x^{\mathscr{A}} = \nu_x^{\mathscr{A}}$ for all $x \in Y'$.

1.7. **Countably generated sigma-subalgebras.** Now let $\mathscr{A}$ be a countably generated σ-subalgebra of $\mathcal{B}_X$. For $x \in X$, define

$$[x]_{\mathscr{A}} := \bigcap_{x \in A \in \mathscr{A}} A.$$

We sometimes refer to $[x]_{\mathscr{A}}$ as the **atom** containing x.

Lemma 1.13. *Take $(A_1, A_2, ...)$ be such that $\mathscr{A}$ is the smallest σ-subalgebra of $\mathcal{B}_X$ containing all A_i's. Fix $x \in X$ and let*

$$B_i := \begin{cases} A_i, & \text{if } x \in A_i \\ A_i^c, & \text{if } x \notin A_i. \end{cases}$$

Then $[x]_{\mathscr{A}} = \bigcap_{i=1}^{\infty} B_i$. In particular, $[x]_{\mathscr{A}} \in \mathscr{A}$.

Proof. Fix some $x \in X$ and let $[x]' := \bigcap_{i=1}^{\infty} B_i$. Consider

$$\mathscr{A}'_x := \{A \in \mathscr{A} \mid \text{either } [x]' \subset A \text{ or } [x]' \subset A^c\}$$

Then one verifies that $\mathscr{A}'_x$ is a σ-algebra containing all A_i's. Thus it is equal to $\mathscr{A}$. This proves the lemma. ☐

Fix a countable generator $(A_1, A_2, ...)$ of $\mathscr{A}$. Let $(B_1, B_2, ...)$ be obtained by including their complements.

By part (2) of the theorem, we can find $X_1 \in \mathscr{A}$ contained in X' (defined by intersection of all X'_{B_i}'s) of full measure such that

$$\int_{A \cap X_1} \left(\int_X \mathbf{1}_{B_i}(\omega)\, \mu_x^{\mathscr{A}}(\omega) \right) \mu(x) = \int_A \mathbf{1}_{B_i}(\omega)\, \mu(\omega), \quad \forall\, A \in \mathscr{A}.$$

Let $N_i \in \mathscr{A}$ defined by $\left\{x \in B_i \cap X_1 \,\middle|\, \mu_x^{\mathscr{A}}(B_i) \neq 1\right\}$. Then the equation above (with $A := N_i$) shows that $\mu(N_i) = 0$. Let $X'' := X_1 \setminus \bigcup N_i$.

Then for $x \in X''$ and $x \in B_i$, one has $\mu_x^{\mathscr{A}}(B_i) = 1$. Hence $\mu_x^{\mathscr{A}}([x]_{\mathscr{A}}) = 1$.

As for each $f \in \mathscr{C}$, the map $x \mapsto \mu_x^{\mathscr{A}}(f)$ is $\mathscr{A}$-measurable (on X'), it must be constant on each atom. So $\mu_x^{\mathscr{A}} = \mu_y^{\mathscr{A}}$ whenever $[x]_{\mathscr{A}} = [y]_{\mathscr{A}} \subset X'$.

1.8. **Pointwise convergence.** Choose $X' \subset X$ such that for $\star = \mathscr{A}_i$ or $\mathscr{A}_\infty$, conditional measures $\mu_x^\star$ are defined and (1) and (2) in this theorem hold.

By (2) of the theorem, for every $E \in \mathscr{A}_\infty$, (the equivalence class of) the function $x \mapsto \mu_x^\star(E)$ is just $\pi_\star([\mathbf{1}_E])$ for $\star = \mathscr{A}_?$ for $? = 1, 2, ...$ or ∞.

First note that

Lemma 1.14. *$\bigcup_{i=1}^{\infty} L^2(X, \mathscr{A}_i, \mu)$ is dense in $L^2(X, \mathscr{A}_\infty, \mu)$.*

Proof. It suffices to show that for every $E \in \mathscr{A}_\infty$ and $\varepsilon > 0$, there exists $F \in \bigcup_{i=1}^{\infty} \mathscr{A}_i$ such that $\mu(E \Delta F) < \varepsilon$. Let $\mathscr{A}'$ collect all $E \in \mathscr{A}_\infty$ such that this holds, then one checks that $\mathscr{A}'$ is a σ-subalgebra containing all the $\mathscr{A}_i$'s. Consequently, $\mathscr{A}' = \mathscr{A}_\infty$ and the proof is complete. ☐

Thus the convergence in L^2-norm is guaranteed by general facts from Hilbert spaces. However, going from L^2-convergence to pointwise convergence requiring us to pay attention to the special choices of subspaces here (think about Carleson's difficult theorem on pointwise convergence of Fourier series). The key is the following "maximal inequalities":

Lemma 1.15. *For $f \in L^1(X, \mathscr{B}, \mu)$ (in application, $f = \mathbf{1}_E - \mathbf{1}_F$) and $\lambda > 0$, let*

$$E(\lambda) := \left\{ x \in X' \,\middle|\, \max_{n \in \mathbb{Z}^+} \mu_x^{\mathscr{A}_n}(f) > \lambda \right\}$$

Then

$$\mu\left(E(\lambda)\right) \leq \lambda^{-1} \|f\|_1 .$$

Proof. If all $\mathscr{A}_i$'s are the same, this is just Minkowski inequality. In general, let

$$F_1 := \left\{x \in X' \,\middle|\, \mu_x^{\mathscr{A}_1}(f) > \lambda\right\} \in \mathscr{A}_1;$$
$$F_2 := \left\{x \in X' \setminus F_1 \,\middle|\, \mu_x^{\mathscr{A}_2}(f) > \lambda\right\} \in \mathscr{A}_2;$$
$$F_3 := \left\{x \in X' \setminus (F_1 \cup F_2) \,\middle|\, \mu_x^{\mathscr{A}_3}(f) > \lambda\right\} \in \mathscr{A}_3;$$

......

Then $E(\lambda) = \bigsqcup_{k=1}^{\infty} F_k$. For every $k \in \mathbb{Z}^+$,

$$\lambda\mu(F_k) \le \int_{F_k} \mu_x^{\mathscr{A}_k}(f)\mu(x) = \int_{F_k} f(\omega)\mu(\omega) \le \int_{F_k} |f(\omega)|\,\mu(\omega).$$

Thus,

$$\mu(E(\lambda)) = \sum \mu(F_k) \le \lambda^{-1} \sum \int_{F_k} |f(\omega)|\,\mu(\omega) \le \lambda^{-1}\,\|f\|_1\,.$$

□

Proof of (5) of Theorem 1.1. Fix $E \in \mathscr{A}_\infty$ and we would like to show that $\mu_x^{\mathscr{A}_n}(E)$ converges to $\mu_x^{\mathscr{A}_\infty}(E)$ almost surely. So for $\varepsilon > 0$, let

$$E(\varepsilon) := \left\{x \;\middle|\; \limsup \left|\mu_x^{\mathscr{A}_n}(E) - \mu_x^{\mathscr{A}_\infty}(E)\right| > \varepsilon\right\}.$$

It suffices to show that $\mu(E(\varepsilon)) \le 4\varepsilon$ for every $\varepsilon > 0$.

Take $k = k(\varepsilon) \in \mathbb{Z}^+$ and $F \in \mathscr{A}_k$ such that

$$\|\mathbf{1}_E - \mathbf{1}_F\|_2 < \varepsilon^2.$$

Note that

$$\limsup \left|\mu_x^{\mathscr{A}_n}(E) - \mu_x^{\mathscr{A}_\infty}(E)\right| \le \limsup \left|\mu_x^{\mathscr{A}_n}(E) - \mu_x^{\mathscr{A}_n}(F)\right| + \limsup \left|\mu_x^{\mathscr{A}_n}(F) - \mu_x^{\mathscr{A}_\infty}(E)\right|.$$

And for n larger than k, $\mu_x^{\mathscr{A}_n}(F) = \mu_x^{\mathscr{A}_\infty}(F) = \mathbf{1}_F(x)$ almost surely. So $E(\varepsilon) \subset F(\varepsilon) \cup G(\varepsilon)$ where

$$F(\varepsilon) := \left\{x \;\middle|\; \limsup_n \left|\mu_x^{\mathscr{A}_n}(E) - \mu_x^{\mathscr{A}_n}(F)\right| > 0.5\varepsilon\right\},$$
$$G(\varepsilon) := \left\{x \;\middle|\; \left|\mu_x^{\mathscr{A}_\infty}(F) - \mu_x^{\mathscr{A}_\infty}(E)\right| > 0.5\varepsilon\right\}.$$

By Lemma 1.15, we have

$$\mu(F(\varepsilon)),\ \mu(G(\varepsilon)) \le \frac{2}{\varepsilon}\,\|\mathbf{1}_E - \mathbf{1}_F\|_1 \le 2\varepsilon.$$

Hence $\mu(E(\varepsilon)) \le 4\varepsilon$ as desired. □

References

[EW11] Manfred Einsiedler and Thomas Ward, *Ergodic theory with a view towards number theory*, Graduate Texts in Mathematics, vol. 259, Springer-Verlag London, Ltd., London, 2011. MR 2723325 1

[Rud87] Walter Rudin, *Real and complex analysis*, third ed., McGraw-Hill Book Co., New York, 1987. MR 924157 3