Lecture Note for MAT8030: Advanced Probability

LI Liying

September 10, 2024

1 Measure theory preliminaries

In this section we cover some basic facts in measure theory and how they integrate into the modern probability theory, which is essential to this field. Most of the materials are still within the scope of the celebrated work, *Foundations of the theory of probability*, by Kolmogorov in 1933 ([Kol33]).

1.1 Random variables, σ-fields and measures

We start with examples of some random variables (r.v.'s) that the reader should be familiar with from elementary probability. There are two types of r.v.'s encountered in elementary probability: discrete and continuous.

Example 1.1 Examples of discrete r.v.'s.

- **Bernoulli:** $X \sim \mathrm{Ber}(p)$, with $\mathsf{P}(X=1)=p$, $\mathsf{P}(X=0)=1-p$.
- **binomial:** $X \sim \mathrm{Binom}(n,p)$ with $\mathsf{P}(X=k)=\binom{n}{k}p^k(1-p)^{n-k}$, $k=0,1,\dots,n$.
- **geometry:** $X \sim \mathrm{Geo}(p)$, with $\mathsf{P}(X=k)=(1-p)^{k-1}p$, $k=1,2,\dots$.
- **Poisson:** $X \sim \mathrm{Poi}(\lambda)$, with $\mathsf{P}(X=k)=\frac{\lambda^k}{k!}e^{-\lambda}$, $k=0,1,\dots$.

Example 1.2 Examples of continuous r.v.'s, described by the density function $\mathsf{P}(X\le a)=\int_{-\infty}^a p(x)\,dx$.

- **exponential:** $X \sim \mathrm{Exp}(\lambda)$, with $p(x)=\mathbb{1}_{[0,\infty)}(x)\cdot\lambda e^{-\lambda x}$.
- **uniform:** $X \sim \mathrm{Unif}[a,b]$, with $p(x)=\mathbb{1}_{[a,b]}(x)\cdot\frac{1}{b-a}$.
- **normal/Gaussian:** $X \sim \mathcal{N}(\mu,\sigma^2)$, with $p(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-(x-\mu)^2/2\sigma^2}$.

Recall that the distribution/law of a r.v. X is determined by its cumulative distribution function (c.d.f.). In particular, sets of the form $\{X\le a\}$ are *events* of which one can evaluate the probability, denoted by $\mathsf{P}(X\le a)$.

We can say that $\mathsf{P}(\cdot)$ is a function of events, or a *set function*. A measure $\mathsf{P}(\cdot): A\mapsto \mathsf{P}(A)\in[0,\infty)$ is a special set function satisfying the following three properties:

1. **non-negativity**: $\mathsf{P}(A)\ge 0$, $\forall A$.
2. $\mathsf{P}(\varnothing)=0$.
3. **countable additivity**: for any *disjoint* $A_1,A_2,\dots$,

$$\mathsf{P}\Big(\bigcup_{n=1}^{\infty}A_n\Big)=\sum_{n=1}^{\infty}\mathsf{P}(A_n). \tag{1.1}$$

The last property, countable additivity (a.k.a. σ-additivity) is the most important one. It is only with σ-additivity, not finite additivity, that one can get the hands on various limit theorems for integration/expectation.

Other important properties of measures can be derived from Item 1 to Item 3.

4. **finite additivity** from Items 2 and 3: let $A_{n+1} = A_{n+2} = \cdots = \varnothing$ in (1.1); then
$$\mathsf{P}\Big(\bigcup_{k=1}^{n} A_k\Big) = \sum_{k=1}^{n} \mathsf{P}(A_k).$$
5. **monotonicity** from Items 1 and 4: if $A \subset B$, then $A \cap (B \setminus A) = \varnothing$, and hence
$$\mathsf{P}(B) = \mathsf{P}(A) + \mathsf{P}(B \setminus A) \geq \mathsf{P}(A).$$
6. **sub-additivity** from Items 3 and 5: let $\tilde{A}_n = A_n \setminus \big(\bigcup_{k=1}^{n-1} A_k\big) \subset A_n$; then
$$\mathsf{P}\Big(\bigcup_{n=1}^{\infty} A_n\Big) = \sum_{n=1}^{\infty} \mathsf{P}(\tilde{A}_n) \leq \sum_{n=1}^{\infty} \mathsf{P}(A_n).$$
7. **continuity from above** from Items 2 and 3: if $A_n \downarrow A$ and $\mathsf{P}(A_1) < \infty$, then $\mathsf{P}(A) = \lim_{n\to\infty} \mathsf{P}(A_n)$ ($A = \bigcap_{n=1}^{\infty} A_n$). In fact, since A_1 is the disjoint union of
$$A_1 = A \cup (A_1 \setminus A_2) \cup (A_2 \setminus A_3) \cup \cdots, \tag{1.2}$$
we have
$$\mathsf{P}(A_1) = \mathsf{P}(A) + \mathsf{P}(A \setminus A_n) + \sum_{k=n}^{\infty} \mathsf{P}(A_k \setminus A_{k+1}).$$
All the terms are positive, and the left hand side is finite, so the tail of the infinite sum must converges to 0, and hence
$$\mathsf{P}(A) = \lim_{n\to\infty} \mathsf{P}(A_1) - \mathsf{P}(A \setminus A_n) - \sum_{k=n}^{\infty} \mathsf{P}(A_k \setminus A_{k+1}) = \lim_{n\to\infty} \mathsf{P}(A_1) - \mathsf{P}(A_1 \setminus A_n) = \lim_{n\to\infty} \mathsf{P}(A_n).$$
Note: the decomposition (1.2) *has the following interpretation; as A_n is decreasing, any element $x \in A_1$ either appears in all A_n, and hence in A, or there is a largest n such that $x \in A_n$ but $x \notin A_{n+1}$, and hence $x \in A_n \setminus A_{n+1}$.*
8. **continuity from below** from Items 2, 3, 5 and 7: if $A_n \uparrow A$, then $\mathsf{P}(A) = \lim_{n\to\infty} \mathsf{P}(A_n)$.
Noting that $\mathsf{P}(A_n)$ is increasing, by sub-additivity,
$$\mathsf{P}(A) \leq \mathsf{P}(A_1) + \sum_{n=2}^{\infty} \mathsf{P}(A_n \setminus A_{n-1}) = \lim_{n\to\infty} \mathsf{P}(A_n).$$
If $\mathsf{P}(A) = \infty$, there is nothing else to prove. Otherwise, $\mathsf{P}(A) < \infty$, and $A - A_n \downarrow \varnothing$. Then by continuity from above,
$$0 = \mathsf{P}(\varnothing) = \lim_{n\to\infty} \mathsf{P}(A \setminus A_n) = \lim_{n\to\infty} \mathsf{P}(A) - \mathsf{P}(A_n).$$

We also need to impose some conditions on the domain of the set function $\mathsf{P}(\cdot)$. The domain should behave well under countable union/intersection. This leads to the definition of σ-algebras.

Definition 1.1 *Let Ω be any non-empty set and $\mathcal{F}$ be a collection of subsets of Ω. We say that $\mathcal{F}$ is a σ-algebra (or σ-field), if*

1. $\Omega \in \mathcal{F}$,
2. $A \in \mathcal{F}$ *implies* $A^c \in \mathcal{F}$,
3. *(closure under countable union)* $A_n \in \mathcal{F}$ *implies* $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$.

Example 1.3 1. The smallest σ-algebra: $\mathcal{F} = \{\varnothing, \Omega\}$.

2. The largest σ-algebra: $\mathcal{F} = \{$ all subsets of $\Omega\}$.

A set Ω equipped with a σ-algebra $\mathcal{F}$ is called a *measurable space*, written in a pair $(\Omega, \mathcal{F})$.

Proposition 1.1 [1] *Let $\mathcal{F}$ be a σ-field. Then*

- $\varnothing \in \mathcal{F}$,
- $A \subset B$, $A, B \in \mathcal{F}$ *imply* $B \setminus A \in \mathcal{F}$,
- *(closure under countable intersection)* $A_n \in \mathcal{F}$ *implies* $\bigcap_{n=1}^{\infty} A_n \in \mathcal{F}$.

Definition 1.2 *A probability space, or probability triple, $(\Omega, \mathcal{F}, \mathsf{P})$ is such that $(\Omega, \mathcal{F})$ is a measurable space and $\mathsf{P} : \mathcal{F} \to [0, 1]$ is a measure with $\mathsf{P}(\Omega) = 1$.*

Definition 1.3 *A random variable (r.v.) $X = X(\omega) : \Omega \to \mathbb{R}$ is a map from a probability space $(\Omega, \mathcal{F}, \mathsf{P})$ to $\mathbb{R}$, such that*

$$\{\omega : X(\omega) \leq a\} \in \mathcal{F}, \quad \forall a \in \mathbb{R},$$

or written more compactly, $X^{-1}\big((-\infty, a]\big) \in \mathcal{F}$ for all $a \in \mathbb{R}$.

Let us recall some basic facts about the pre-image map φ^{-1} for *any* map $\varphi : U \to V$. It is defined by

$$\varphi^{-1}(W) := \{u \in U : \varphi(u) \in W\}.$$

Proposition 1.2 *The map φ^{-1} commutes with most set operations, in particular:*

- $\varphi^{-1}(W_1 \cap W_2) = \varphi^{-1}(W_1) \cap \varphi^{-1}(W_2)$,
- $\varphi^{-1}(W_1 \cup W_2) = \varphi^{-1}(W_1) \cup \varphi^{-1}(W_2)$,
- $\varphi^{-1}(W^c) = \big(\varphi^{-1}(W)\big)^c$.

Let X be a r.v. on $(\Omega, \mathcal{F}, \mathsf{P})$, and let $\mathcal{B} = \{A \text{ s.t. } X^{-1}(A) \in \mathcal{F}\}$. Definition 1.3 and Proposition 1.2 imply that $\mathcal{B}$ contains all the intervals in $\mathbb{R}$. Moreover, since $\mathcal{F}$ is a σ-algebra,

$$X^{-1}(I_n) \in \mathcal{F} \quad \Longrightarrow \quad X^{-1}\Big(\bigcup_{n=1}^{\infty} I_n\Big) = \bigcup_{n=1}^{\infty} X^{-1}(I_n) \in \mathcal{F}.$$

This implies that $\mathcal{B}$ is also a σ-algebra. As we will see in the next section, $\mathcal{B}$ contains the *Borel σ-algebra*, which is the most important class of σ-algebras in probability theory.

[1] In this note, readers are encouraged to work out their own proofs on propositions without proofs; they are good exercises and will be useful for understanding later materials.

1.2 Construction of σ-algebra and (probability) measures

Simply put, the Borel σ-algebra is the *smallest* σ-algebra containing by open sets. To understand what is "smallest", we start with the following observation.

Lemma 1.3 *1. If $\mathcal{F}_1$ and $\mathcal{F}_2$ are two σ-algebras on Ω, then $\mathcal{F}_1 \cap \mathcal{F}_2$ is also a σ-algebra.*

2. If $\mathcal{F}_\gamma, \gamma \in \Gamma$ are σ-algebras on Ω, where Γ is an arbitrary index set (countable or uncountable), then $\bigcap_{\gamma\in\Gamma} \mathcal{F}_\gamma$ is also a σ-algebra.

Proposition 1.4 *Let $\mathcal{A}$ be a collection of subsets in Ω. Then there exists a smallest σ-algebra containing $\mathcal{A}$, called* the σ-algebra generated by $\mathcal{A}$ *and written $\sigma(\mathcal{A})$, in the sense that if $\mathcal{G} \supset \mathcal{A}$ is a σ-algebra, then $\sigma(\mathcal{A}) \subset \mathcal{G}$.*

Proof: Take $\sigma(\mathcal{A}) = \bigcap_{\mathcal{F}\ \sigma\text{-algebra}:\mathcal{F}\supset\mathcal{A}} \mathcal{F}$. □

Definition 1.4 (Borel σ-algebra) *Let M be a metric space (or any topological space). The* Borel σ-algebra $\mathcal{B}(M)$ *is the σ-algebra generated by all the open sets in M.*

Example 1.4

- $\mathcal{B}(\mathbb{R}) = \sigma\Big((-\infty, a], a \in \mathbb{R}\Big)$.
- $\mathcal{B}(\mathbb{R}^d) = \sigma\Big((-\infty, a_1] \times \cdots \times (-\infty, a_d], a_i \in \mathbb{R}\Big)$.

Remark 1.5 Here, one need to first show that any open sets in $\mathbb{R}^d$ can be obtained from countable union of sets of the form $(-\infty, a_1] \times (-\infty, a_d]$. The construction requires some ideas from point-set topology, but it is elementary, and thus omitted here.

Proposition 1.5 *A map $X(\omega)$ on $(\Omega, \mathcal{F}, \mathsf{P})$ is a r.v. if and only if $X^{-1}(A) \in \mathcal{F}$ for any $A \in \mathcal{B}(\mathbb{R})$.*

Remark 1.6 In fact, this is usually taken as the definition for r.v.'s.

Now let us take about the distribution of a r.v. X. One can check that $\mu = \mathsf{P} \circ X^{-1}$ defined by

$$\mu(A) = \mathsf{P}(\{\omega : X(\omega) \in A\}), \quad A \in \mathcal{B}(\mathbb{R}),$$

is a probability measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. We call μ the *distribution/law* of X. Clearly, $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mu)$ is a probability space. For most of the practical application, say computing expectation, variance, etc, it is enough to understand the distribution of a r.v., not the original probability measure P on some abstract space that can be potentially be very complicate. Another obvious advantage is that the distributions of all r.v.'s are probability measures live on the *same* measurable space $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

Note that the c.d.f. of a r.v. can be read from its distribution:

$$F_X(a) = \mathsf{P}(X \le a) = \mu\big((-\infty, a]\big), \quad a \in \mathbb{R}.$$

The central topic for this section is to understand how the c.d.f. determines μ. Along the way we will learn how to construct σ-algebras and (probability) measures. Some of the presentation here is taken from [Shi96, Chap. 2.3]. The next theorem is a fundamental and important result.

Theorem 1.6 *Every increasing, right-continuous function $F : \mathbb{R} \to [0, 1]$ with $F(-\infty) = 0$ and $F(\infty) = 1$ uniquely determines a probability measure μ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.*

We start by introducing some notions on collections of sets.

Definition 1.5 *A collection of sets $\mathcal{S}$ is a* semi-algebra *if first, it is closed under intersection, i.e., $A, B \in \mathcal{S} \Rightarrow A \cap B \in \mathcal{S}$ and second, for every $A \in \mathcal{S}$, A^c is disjoint union of $A_1, A_2, \ldots, A_n$ in $\mathcal{S}$.*

A collection of sets $\mathcal{S}$ is an algebra, *or* field, *if $A, B \in \mathcal{S}$ implies $A \cap B \in \mathcal{S}$ and $A^c \in \mathcal{S}$.*

These two notions are related by the following proposition.

Proposition 1.7 *Let $\mathcal{S}$ be a semi-algebra. Then*

$$\bar{\mathcal{S}} = \{\textit{finite disjoint unions of sets in } \mathcal{S}\}$$

is an algebra.

Example 1.7 All the d-dimensional half-open, half-closed rectangles forms a semi-algebra:

$$\mathcal{S}_d = \{\varnothing, \quad (a_1, b_1] \times \cdots \times (a_d, b_d], \ -\infty \le a_i < b_i \le \infty\}.$$

Definition 1.6 *A collection of sets $\mathcal{S}$ is a* monotone class (m-class), *if for every monotone sequence $A_n \in \mathcal{S}$, $A = \lim_{n\to\infty} A_n \in \mathcal{S}$.*

Here, for an increasing sequence $A_n \subset A_{n+1} \subset \cdots$, its limit is defined by $A := \bigcup_{n=1}^{\infty} A_n$, and for an decreasing sequence $A_n \supset A_{n+1} \supset \cdots$, its limit is defined by $A := \bigcap_{n=1}^{\infty} A_n$.

It is easy to see that any *intersection* of m-classes is still an m-class. Therefore, it makes sense to talk about the *smallest* m-classes containing any collection of sets $\mathcal{A}$ (c.f. Proposition 1.4). We denote this smallest m-class by $m(\mathcal{A})$.

The monotone class condition basically bridges the difference between σ-algebras and algebras.

Proposition 1.8 *Let $\mathcal{A}$ be a collection of subsets of Ω. Then $\mathcal{A}$ is a σ-algebra if and only if $\mathcal{A}$ is both an algebra and an m-class.*

Theorem 1.9 *(Monotone Class Theorem) Let $\mathcal{A}$ be an algebra. Then $\sigma(\mathcal{A}) = m(\mathcal{A})$.*

Proof: By Proposition 1.8, $\sigma(\mathcal{A})$ is necessarily an m-class, and by the minimum property we have the inclusion $m(\mathcal{A}) \subset \sigma(\mathcal{A})$.

To show the other direction $\sigma(\mathcal{A}) \subset m(\mathcal{A})$, it suffices to show that $m(\mathcal{A})$ is an algebra, and hence a σ-algebra (using Proposition 1.8 again). To establish that $m(\mathcal{A})$ is an algebra, we will use the *principle of appropriate sets.*

First, $m(\mathcal{A})$ is closed under complement. Let

$$\mathcal{S} = \{A : A, A^c \in m(\mathcal{A})\} \subset m(\mathcal{A}).$$

Our goal is to show that $m(\mathcal{A}) = \mathcal{S}$. Clearly, by definition we have $\mathcal{A} \in \mathcal{S}$. Moreover, $\mathcal{S}$ is an m-class: if $A_n \uparrow A$, $A_n \in \mathcal{S}$, then A_n, A_n^c are both monotone sequence in $m(\mathcal{A})$, and hence their limits $A, A^c \in m(\mathcal{A})$; if $A_n \downarrow A$ it is similar. Therefore, $\mathcal{S}$ must contain the smallest m-class that contains $\mathcal{A}$, which is $m(\mathcal{A})$. This shows $\mathcal{S} = m(\mathcal{A})$, so by the definition of $\mathcal{S}$, $m(\mathcal{A})$ is closed under complement.

Second, $m(\mathcal{A})$ is closed under intersection. Since intersection involves two sets, the proof is slightly more complicated and we will do it in two steps. In the first step, for fixed $A \in \mathcal{A}$, let

$$\mathcal{S}_A = \{B : B \in m(\mathcal{A}), \ A \cap B \in m(\mathcal{A})\} \subset m(\mathcal{A}).$$

Clearly, $\mathcal{A} \subset \mathcal{S}_A$ since A is an algebra and $m(\mathcal{A})$ contains $\mathcal{A}$. Also, $\mathcal{S}_A$ is an m-class as $B_n \downarrow B$ or $B_n \uparrow B$ implies $A \cap B_n \downarrow A \cap B$ or $A \cup B_n \uparrow A \cup B$. Therefore, $m(\mathcal{A}) \subset \mathcal{S}_A$, and we have shown that $A \cap B \in m(\mathcal{A})$ whenever $A \in \mathcal{A}$ and $B \in m(\mathcal{A})$.

In the second step, let

$$\mathcal{S} = \{A \in m(\mathcal{A}) : A \cap B \in m(\mathcal{A}),\ \forall B \in m(\mathcal{A})\}.$$

By the first step, $\mathcal{A} \subset \mathcal{S}$. Again, it is not hard to check that A is an m-class. Therefore $m(\mathcal{A}) = \mathcal{S}$, and this proves that $m(\mathcal{A})$ is closed under intersection.

In conclusion, $m(\mathcal{A})$ is an algebra and hence a σ-algebra, this completes the proof. $\square$

A related concept is the Dynkin system (d-system, λ-class).

Definition 1.7 *Let $\mathcal{D}$ be a collection of subsets of Ω. We say that $\mathcal{D}$ is a Dynkin system if*

1. $\Omega \in \mathcal{D}$,
2. $A, B \in \mathcal{D},\ A \subset B \quad \Rightarrow \quad B \setminus A \in \mathcal{D}$,
3. $A_n \uparrow A,\ A_n \in \mathcal{D} \quad \Rightarrow \quad A \in \mathcal{D}$.

We say that $\mathcal{A}$ is a π-system if it is closed under intersection. One can check that $\mathcal{A}$ is a σ-algebra if and only if it is both a π-system and Dynkin system. Moreover, analogous to Theorem 1.9, the following is true.

Theorem 1.10 *(π-λ Theorem; Dynkin Theorem) If $\mathcal{A}$ is a π-system, then $\sigma(\mathcal{A})$ is the smallest Dynkin system containing $\mathcal{A}$.*

Proof: The proof can be done via the principle of appropriate sets. $\square$

Given a distribution function F as in Theorem 1.6, we can introduce a (probability) measure μ_0 on the algebra

$$\bar{\mathcal{S}} = \{\bigcup_{k=1}^{n} (a_k, b_k],\ \text{disjoint union}\},$$

given by

$$\mu_0(A) = \sum_{k=1}^{n} \big[F(b_k) - F(a_k)\big].$$

It is easy to check that μ_0 is finitely additive. An important step is the following.

Proposition 1.11 *The finitely additive measure μ_0 is σ-additive on $\bar{\mathcal{S}}$, i.e., if $A_n \in \bar{\mathcal{S}}$ are disjoint and $\bigcup_{n=1}^{\infty} A_n \in \bar{\mathcal{S}}$, then*

$$\mu_0(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mu_0(A_n).$$

Proof: We will use the fact that σ-additive is equivalent to continuity at $\varnothing$, i.e., μ_0 is σ-additive if and only if for every $A_n \downarrow \varnothing$, $\lim_{n\to\infty} \mu_0(A_n) = \mu_0(\varnothing) = 0$.

Suppose that there is some $L > 0$ such that $A_n \in [-L, L]$. Let $\varepsilon > 0$. By the right continuity of F, there exists $B_n \in \bar{\mathcal{S}}$ such that $\overline{B_n} \subset A_n$ and

$$\mu_0(A_n) - \mu_0(B_n) \le \varepsilon \cdot 2^{-n}.$$

(To see this, write $A_n = \bigcup_{i=1}^{m} (a_i^{(n)}, b_i^{(n)}]$, and let $B_n = \bigcup_{i=1}^{m} (a_i^{(n)} + \delta, b_i^{(n)}]$. Then

$$\mu_0(A_n) - \mu(B_n) = \sum_{i=1}^{m} \big(F(b_i^{(n)} + \delta) - F(b_i^{(n)})\big) \to 0, \quad \delta \downarrow 0.$$

We just need to choose δ small enough so that the sum is less that $\varepsilon \cdot 2^{-n}$.)

Since $A_n \downarrow \varnothing$ and $\overline{B_n} \subset A_n$, we have $\overline{B_n} \downarrow \varnothing$. So $C_n = [-L, L] \setminus \overline{B_n}$ forms an open cover of $[-L, L]$. By the Finite Open Cover Theorem, there is a finite sub-cover, i.e., $\exists n_0$ s.t.

$$[-L, L] \subset \bigcup_{n=1}^{n_0} [-L, L] \setminus \overline{B_n},$$

and hence $\bigcap_{n=1}^{n_0} \overline{B_n} = \varnothing$. Therefore,

$$\mu_0(A_{n_0}) = \mu_0\big(A_{n_0} \setminus \bigcap_{n=1}^{n_0} B_n\big) \le \mu_0\Big(\bigcup_{n=1}^{n_0} (A_n \setminus B_n)\Big) \le \sum_{n=1}^{n_0} \mu_0(A_n \setminus B_n) \le \varepsilon \sum_{n=1}^{\infty} 2^{-n} \le \varepsilon.$$

Noting that $\mu_0(A_n)$ is decreasing and ε is arbitrary, we have $\lim_{n\to\infty} \mu_0(A_n) = 0$.

For unbounded A_n, since $F(-\infty) = 0$ and $F(\infty) = 1$, for every $\varepsilon > 0$, we can choose L s.t. $\mu_0((-L, L]) \ge 1 - \varepsilon$. Let $\tilde{A}_n = A_n \cap (-L, L]$. Then $\tilde{A}_n \downarrow \varnothing$ and $\tilde{A}_n$ are bounded. Hence, $\lim_{n\to\infty} \mu_0(\tilde{A}_n) = 0$. Therefore,

$$\limsup_{n\to\infty} \mu_0(A_n) \le \limsup_{n\to\infty} \mu_0(\tilde{A}_n) + \limsup_{n\to\infty} \mu_0\big(A_n \setminus (-L, L]\big) \le 0 + \varepsilon = \varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, we see that $\lim_{n\to\infty} \mu_0(A_n) = 0$, as desired. $\square$

References

[Kol33] A.N. Kolmogorov. *Foundations of the Theory of Probability (English Translation)*. 1933.

[Shi96] A. N. Shiryaev. *Probability*, volume 95 of *Graduate Texts in Mathematics*. Springer New York, 1996.