Lecture 2

Instructor: Quan Zhou

For more details about the materials covered in this note, see Chapters 2.1, 2.2 and 2.4 of Resnick [3] and Chapter 1.1 and Appendix A of Durrett [2].

2.1 Measures and measure spaces

Definition 2.1. Given a measurable space $(\Omega, \mathcal{F})$, a function $\mu : \mathcal{F} \to [0, \infty]$ is a measure if

- $\mu(A) \geq 0$ for any $A \in \mathcal{F}$;
- $\mu(\emptyset) = 0$;
- if $\{A_1, A_2, \dots\}$ is a countable sequence of disjoint sets in $\mathcal{F}$, then $\mu(\cup_i A_i) = \sum_i \mu(A_i)$. This is called countable additivity (or σ-additivity).

$(\Omega, \mathcal{F}, \mu)$ is called a measure space, and sets in $\mathcal{F}$ are called measurable sets. If $\mu(\Omega) = 1$, we call μ a probability measure and $(\Omega, \mathcal{F}, \mu)$ a probability space (or a probability triple).

Remark 2.1. For convenience, we will often deal with the extended real line $\bar{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$. The arithemetic operations involving $\pm\infty$ are defined as follows: (1) $a \pm \infty = \pm\infty$ for any $a \in \mathbb{R}$; (2) $\infty + \infty = \infty$; (3) $a \cdot \infty = \infty$ for any $a \in (0, \infty)$; (4) $\infty \cdot \infty = \infty$. Note that $\infty - \infty$, $0 \cdot \infty$ and ∞/∞ are not defined. In measure theory, it is usually fine to assume that $0 \cdot \infty = 0$ but a rigorous proof is always preferred.[1]

Example 2.1. The following examples are important for probability theory.

(i) Let Ω be a discrete sample space (finite or countably infinite). The counting measure on $(\Omega, \mathcal{P}(\Omega))$ is denoted by $\#$. For any $A \in \mathcal{P}(\Omega)$, $\#(A)$ is equal to the number of elements in A.

[1] An example we will see later is the Lebesgue integral $\int_A f \, d\mu$ with $\mu(A) = 0$ and $f \in [0, \infty]$. One can use the definition of Lebesgue integrals to rigorous prove that $\int_A f \, d\mu = 0$. This justifies a seemingly simpler argument: $\int_A f \, d\mu \leq \mu(A) \sup f = 0 \cdot \infty = 0$, which is not rigorous in the last step.

(ii) The Lebesgue measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, denoted by m, is given by $m((a, b)) = b - a$ for any $-\infty < a \le b < \infty$ [3, §2.5.1].

(iii) Unit point mass measures (Dirac measures): Given a measurable space $(\Omega, \mathcal{F})$ and some $x \in \Omega$, we can define the Dirac measure at x by $\delta_x(A) = \mathbb{1}_A(x)$ for any $A \in \mathcal{F}$.

(iv) An arbitrary discrete probability measure: Assume $\Omega = \{\omega_1, \omega_2, \dots\}$ and let $\{p_i \ge 0\}_{i=1}^{\infty}$ be a sequence of non-negative real numbers such that $\sum p_i = 1$. Then we can define a probability measure P by letting $\mathsf{P}(\{\omega_i\}) = p_i$ and $\mathsf{P}(A) = \sum_{\omega_i \in A} p_i$ for any $A \in \mathcal{P}(\Omega)$. One can check this is a probability measure on $(\Omega, \mathcal{P}(\Omega))$.

2.2 Properties of measures

Proposition 2.1. *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space. Assume that the sets we mention below are all in $\mathcal{F}$.*

(i) Monotonicity: If $A \subset B$, then $\mu(A) \le \mu(B)$.

(ii) Subadditivity: If $A \subset \cup_i A_i$, then $\mu(A) \le \sum_i \mu(A_i)$.

(iii) Continuity from below: If $A_i \uparrow A$, then $\mu(A_i) \uparrow \mu(A)$.

(iv) Continuity from above: If $A_i \downarrow A$ and $\mu(A_1) < \infty$, then $\mu(A_i) \downarrow \mu(A)$.

(v) Inclusion-exclusion formula: If $\mu(A_i) < \infty$ for $i = 1, 2, \dots, n$, then

$$\mu\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{k=1}^{n} \left\{ (-1)^{k-1} \sum_{I \subset \{1,2,\dots,n\}:\ \#(I)=k} \mu\left(\bigcap_{i \in I} A_i\right) \right\}.$$

(vi) If $\mu(\cup_n A_n) < \infty$, then

$$\mu(\liminf_{n\to\infty} A_n) \le \liminf_{n\to\infty} \mu(A_n) \le \limsup_{n\to\infty} \mu(A_n) \le \mu(\limsup_{n\to\infty} A_n).$$

Further, if $A_n \to A$, then $\mu(A_n) \to \mu(A)$.

Proof of part (iii). Let $\{A_n\}$ be an increasing sequence, i.e. $A_1 \subset A_2 \subset \cdots$. Define another sequence of sets $\{B_n\}$ by letting $B_1 = A_1$ and $B_n = A_n \cap A_{n-1}^c$ (this can also be written as $B_n = A_n \setminus A_{n-1}$.) Note that $\cup_{i=1}^{n} B_n = A_n$, which implies that $\cup_{n=1}^{\infty} B_n = \lim_{n\to\infty} A_n = A$. Further, $\{B_n\}$ is a disjoint sequence and thus by the σ-additivity of measures,

$$\mu(A) = \mu(\cup_{n\geq 1} B_n) = \sum_{n\geq 1} \mu(B_n) = \lim_{n\to\infty} \sum_{i=1}^{n} \mu(B_i).$$

The last step follows from the monotone convergence theorem for sequences of real numbers, and note that the limit can be infinity. The rest follows by observing that $\sum_{i=1}^{n} \mu(B_i) = \mu(\cup_{i=1}^{n} B_i) = \mu(A_n)$. □

Proof of part (vi). For a sequence of sets $\{A_n\}$, define $B_n = \sup_{k\geq n} A_k$ and $C_n = \inf_{k\geq n} A_k$. Note that both $\{B_n\}$ and $\{C_n\}$ are monotone sequences and by Proposition 1.3, we have $\liminf_{n\to\infty} A_n = \lim_{n\to\infty} C_n$ and $\limsup_{n\to\infty} A_n = \lim_{n\to\infty} B_n$. Assuming $\mu(B_1) = \mu(\cup_{n\geq 1} A_n) < \infty$, by (i) and (iv),

$$\mu(\limsup_{n\to\infty} A_n) = \lim_{n\to\infty} \mu(B_n) = \limsup_{n\to\infty} \mu(B_n) \geq \limsup_{n\to\infty} \mu(A_n).$$

Similarly, $\mu(\liminf_{n\to\infty} A_n) = \lim_{n\to\infty} \mu(C_n) \leq \liminf_{n\to\infty} \mu(A_n)$ by (i) and (iii). The first claim then follows since lim inf (of a real sequence) cannot be greater than lim sup.

If we further assume that $A_n \to A$, which by definition means that $A = \limsup_{n\to\infty} A_n = \liminf_{n\to\infty} A_n$, then

$$\mu(A) \leq \liminf_{n\to\infty} \mu(A_n) \leq \limsup_{n\to\infty} \mu(A_n) \leq \mu(A).$$

Hence, $\limsup_{n\to\infty} \mu(A_n) = \liminf_{n\to\infty} \mu(A_n)$ and $\mu(A_n) \to \mu(A)$. □

Proof of the remaining part(s). Try it yourself. □

Example 2.2. The inclusion-exclusion formula can be proved by using Venn diagram. The simplest case is given by $\mu(A \cup B) = \mu(A) + \mu(B) - \mu(A \cap B)$.

Example 2.3. Let m be the Lebesgue measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Let $A_n = [n, \infty)$. Then, $m(A_n) = \infty$ for every n, but $m(\lim_{n\to\infty} A_n) = m(\emptyset) = 0$.

2.3 Dynkin's π-λ theorem

Definition 2.2. Let $\mathcal{P}, \mathcal{L}$ be two collections of subsets of Ω.

- $\mathcal{P}$ is called a π-system if it is closed under finite intersections.

- $\mathcal{L}$ is called a λ-system if (i) $\emptyset \in \mathcal{L}$; (ii) $\mathcal{L}$ is closed under complementation; (iii) $\mathcal{L}$ is closed under *countable disjoint* unions.

Lemma 2.1. *If a λ-system is closed under finite intersections (i.e. it is also a π-system), then it is a σ-algebra.*

Proof. Try it yourself. □

Theorem 2.1 (Dynkin's π–λ theorem)**.** *If $\mathcal{P}$ is a π-system and $\mathcal{L}$ is a λ-system and $\mathcal{P} \subset \mathcal{L}$, then $\sigma(\mathcal{P}) \subset \mathcal{L}$.*

Proof. Let $\lambda(\mathcal{P})$ denote the minimal λ-system generated by $\mathcal{P}$, which always exists and is unique.

Step (1). For $A \in \lambda(\mathcal{P})$, define $\mathcal{G}_A = \{B \colon A \cap B \in \lambda(\mathcal{P})\}$. We claim $\mathcal{G}_A$ is a λ-system.

First, since $A \cap \Omega = A \in \lambda(\mathcal{P})$, we have $\Omega \in \mathcal{G}_A$.

Second, suppose $B \in \mathcal{G}_A$ which means $A \cap B \in \lambda(\mathcal{P})$ by the definition of $\mathcal{G}_A$. Note that $A \cap B^c = (A^c \cup B)^c = (A^c \cup (A \cap B))^c$. Since both A^c and $A \cap B$ are in $\lambda(\mathcal{P})$ and they are disjoint, $A^c \cup (A \cap B)$ and its complement are also in $\lambda(\mathcal{P})$. Thus, $B^c \in \mathcal{G}_A$.

Third, if $B_1, \dots, B_n$ are disjoint sets in $\mathcal{G}_A$, then $A \cap (\cup_{i=1}^n B_i) = \cup_{i=1}^n (A \cap B_i)$ is a countable disjoint union of sets in $\lambda(\mathcal{P})$, which is also in $\lambda(\mathcal{P})$. Therefore, $\cup_{i=1}^n B_i \in \mathcal{G}_A$.

Step (2). Next, we prove $\lambda(\mathcal{P})$ is a σ-algebra. By Lemma 2.1, it suffices to show that $\lambda(\mathcal{P})$ is closed under finite intersections; that is, for any $A, B \in \lambda(\mathcal{P})$, we have $A \cap B \in \lambda(\mathcal{P})$.

For any $A, B \in \mathcal{P}$, $A \cap B \in \mathcal{P} \subset \lambda(\mathcal{P})$ since $\mathcal{P}$ is a π-system.

This implies that for any $A \in \mathcal{P}$, we have $\mathcal{P} \subset \mathcal{G}_A$. Because $\lambda(\mathcal{P})$ is the minimal λ-system over $\mathcal{P}$, we have $\lambda(\mathcal{P}) \subset \mathcal{G}_A$. It follows from the definition of $\mathcal{G}_A$ that for any $A \in \mathcal{P}$ and $B \in \lambda(\mathcal{P})$, $A \cap B \in \lambda(\mathcal{P})$.

Interchanging the roles of A and B in the previous conclusion, we obtain that for any $A \in \lambda(\mathcal{P})$ and $B \in \mathcal{P}$, $A \cap B \in \lambda(\mathcal{P})$. But this just means $\mathcal{P} \subset \mathcal{G}_A$. Hence, $\lambda(\mathcal{P}) \subset \mathcal{G}_A$ for any $A \in \lambda(\mathcal{P})$, which implies that $\lambda(\mathcal{P})$ is a π-system.

Step (3). By definition, $\sigma(\mathcal{P}) \subset \lambda(\mathcal{P})$ and $\lambda(\mathcal{P}) \subset \mathcal{L}$. Thus, $\sigma(\mathcal{P}) \subset \mathcal{L}$.

The proof is complete. □

Corollary 2.1. *If $\mathcal{P}$ is a π-system, then $\sigma(\mathcal{P}) = \lambda(\mathcal{P})$, where $\lambda(\mathcal{P})$ denotes the minimal λ-system that contains $\mathcal{P}$.*

Proof. Try it yourself. □

Theorem 2.2. *Let P_1, P_2 be two probability measures on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ such that for any $x \in \mathbb{R}$, we have $\mathsf{P}_1((-\infty, x]) = \mathsf{P}_2((-\infty, x])$. Then $\mathsf{P}_1 = \mathsf{P}_2$ on $\mathcal{B}(\mathbb{R})$.*

Proof. This is a very deep result. It tells us the distribution function (which will be defined shortly) uniquely defines a probability measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. We prove the result using Dynkin's theorem.

Step (1). Let $\mathcal{P} = \{(-\infty, x] : x \in \mathbb{R}\}$. Then $\mathcal{P}$ is a π-system since $(-\infty, a] \cap (-\infty, b] = (-\infty, a \wedge b]$.

Step (2). Consider the collection of sets $\mathcal{L} = \{A \in \mathcal{B}(\mathbb{R}) : \mathsf{P}_1(A) = \mathsf{P}_2(A)\}$. Using the properties of probability measures, it is easy to verify that $\mathcal{L}$ is a λ-system.

Step (3). Notice that $\mathcal{P} \subset \mathcal{L}$ and thus $\sigma(\mathcal{P}) \subset \mathcal{L}$. Recalling that $\sigma(\mathcal{P}) = \mathcal{B}(\mathbb{R})$, we conclude that $\mathcal{L} \supset \mathcal{B}(\mathbb{R})$, i.e. P_1 and P_2 agree on $\mathcal{B}(\mathbb{R})$.

The proof is complete. □

2.4 Distribution functions

Definition 2.3. A function $F : \mathbb{R} \to [0, 1]$ is called a distribution function if

- F is right continuous, i.e. $\lim_{x_n \downarrow x} F(x_n) = F(x)$ for every x;
- F is non-decreasing;
- $\lim_{x \to \infty} F(x) = 1$ and $\lim_{x \to -\infty} F(x) = 0$.

Definition 2.4. Quantile functions.

(i) Lower quantile function: $F^-(\alpha) = \inf\{x : F(x) \geq \alpha\}$.

(ii) Upper quantile function: $F^+(\alpha) = \sup\{x : F(x) \leq \alpha\}$.

Example 2.4. A uniform distribution on $(0, 1)$ has distribution function $F(x) = x\mathbb{1}_{(0,1)}(x) + \mathbb{1}_{[1,\infty)}(x)$ (note that F is defined on $\mathbb{R}$).

2.5 Construction of uncountable measure spaces

Definition 2.5. A measure space $(\Omega, \mathcal{F}, \mu)$ is called *σ-finite* if there exists a sequence of sets $A_1, A_2, \dots$ in $\mathcal{F}$ such that $\mu(A_i) < \infty$ for each i and $\Omega = \cup_{i=1}^{\infty} A_i$.

Example 2.5. Examples of σ-finite and non-σ-finite measures.

(i) The Lebesgue measure on the real line and the counting measure defined on some countable space are σ-finite.

(ii) Consider a measure space $(\Omega, \mathcal{F}, \mu)$ such that $\mu(\Omega) > 0$. Define another measure ν by

$$\nu(A) = \begin{cases} 0, & \text{if } \mu(A) = 0, \\ \infty, & \text{if } \mu(A) > 0, \end{cases}$$

for any $A \in \mathcal{F}$. One can show that ν is not σ-finite.

Definition 2.6. An algebra (field) on Ω is a collection of subsets of Ω which contains Ω and is closed under complementation and finite unions.

Definition 2.7. A semi-algebra $\mathcal{S}$ on Ω is a collection of subsets of Ω such that (i) $\emptyset, \Omega \in \mathcal{S}$; (ii) $\mathcal{S}$ is closed under finite intersections; (iii) if $A \in \mathcal{S}$, then A^c is a finite disjoint union of sets in $\mathcal{S}$.

Theorem 2.3. *Let $\mathcal{S}$ be a semi-algebra and $\mu : \mathcal{S} \to [0, \infty]$ be a σ-additive (countably additive) function such that $\mu(\emptyset) = 0$. Then μ has a unique extension which is a measure on the algebra generated by $\mathcal{S}$.*

Proof. See the textbook. □

Theorem 2.4 (Caratheodory's extension theorem)**.** *A σ-finite measure μ on an algebra $\mathcal{A}$ has a unique extension which is a measure on $\sigma(\mathcal{A})$.*

Proof. See the textbook. □

Example 2.6. Consider the sample space $\mathbb{R}^d$ and let $\mathcal{S}_d$ be the collection of all rectangles in $\mathbb{R}^d$ including $\emptyset$, i.e.

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

It can be shown that $\mathcal{S}_d$ is a semi-algebra on $\mathbb{R}^d$. When $d = 1$, we can choose an arbitrary distribution function and define $\mathsf{P} \colon \mathcal{S}_1 \to [0, 1]$ by letting $\mathsf{P}((a, b]) = F(b) - F(a)$. Then, P has a unique extension $\bar{\mathsf{P}}$ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ and $\bar{\mathsf{P}}$ is a probability measure. Indeed, for any set $A \in \mathcal{B}(\mathbb{R})$, we have $\bar{\mathsf{P}}(A) = m(\xi_F(A))$ where m denotes the Lebesgue measure and $\xi_F(A) = \{x \in (0, 1] : F^{-}(x) \in A\}$.

References

[1] Dennis D. Cox. *The Theory of Statistics and Its Applications*. Unpublished.

[2] Rick Durrett. *Probability: Theory and Examples*, volume 49. Cambridge university press, 2019.

[3] Sidney Resnick. *A Probability Path*. Springer, 2019.