NOTES FOR PROBABILITY THEORY

AARON NOTES SERIES

Aaron Xia
Department of Electronic Engineering
Tsinghua University
HaiDian District, Peking
muranqz@gmail.com

May 28, 2019

Contents

1 Preface

My major in Electronic Engineering does not put enough focus on the students' mathematical basis. It's somewhat reasonable for some elementary courses like mathematical analysis and linear algebra but that's not accepatble for probability theory and stochastic process for they are the most important two courses for an EE researcher. Therefore, in order to compensate the weak basis of my probability theory and stochastic process, I decide to study comprehensively by myself. The following sections are mostly summarized from [1] and partly cited from other publications which will be presented in the specific position.

2 Basic Measure Theory

2.1 Basic Abstract Algebra

Definition 2.1 (σ-algebra) *A class of sets $\mathcal{A} \subset 2^\Omega$ is called a **σ-algebra** if it fulfils the following three conditions:*

- $\Omega \in \mathcal{A}$
- *$\mathcal{A}$ is closed under complements.*
- *$\mathcal{A}$ is closed under countable unions.*

Definition 2.2 (algebra) *A class of sets $\mathcal{A} \subset 2^\Omega$ is called an **algebra** if it fulfils the following three conditions:*

- $\Omega \in \mathcal{A}$
- *$\mathcal{A}$ is $\backslash$-closed.*
- *$\mathcal{A}$ is $\cup$-closed.*

Theorem 2.1 *A class of sets $\mathcal{A} \subset 2^\Omega$ is an **algebra** iff. the following three conditions hold:*

- $\Omega \in \mathcal{A}$
- *$\mathcal{A}$ is closed under complements.*
- *$\mathcal{A}$ is closed under intersections.*

Definition 2.3 (ring) *A class of sets $\mathcal{A} \subset 2^\Omega$ is called a **ring** if it fulfils the following three conditions:*

- $\emptyset \in \mathcal{A}$
- *$\mathcal{A}$ is $\backslash$-closed.*
- *$\mathcal{A}$ is $\cup$-closed.*

A ring is called a σ-ring if it is also σ-$\cup$-closed.

Definition 2.4 (semiring) *A class of sets $\mathcal{A} \subset 2^\Omega$ is called a **semiring** if it fulfils the following three conditions:*

- $\emptyset \in \mathcal{A}$
- *for any two sets $A, B \in \mathcal{A}$ the difference set $B \backslash A$ is a finite union of mutually disjoint sets in $\mathcal{A}$.*
- *$\mathcal{A}$ is $\cap$-closed.*

Definition 2.5 (λ-system) *A class of sets $\mathcal{A} \subset 2^\Omega$ is called a **λ-system** (or Dynkin's λ-system) if*

- $\Omega \in \mathcal{A}$
- *for any two sets $A, B \in \mathcal{A}$ with $A \subset B$, the difference set $B \backslash A$ is in $\mathcal{A}$*
- $\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$ *for any choice of countably many pairwise disjoint sets $A_1, A_2, \cdots \in \mathcal{A}$*

Definition 2.6 (π-system) *If $\mathcal{A}$ is a class of sets, $\cap$-closed (closed under intersections) or a π-system if $A \cap B \in \mathcal{A}$ whenever $A, B \in \mathcal{A}$*

Theorem 2.2 (Generated σ-algebra) *Let $\mathcal{E} \subset 2^{\Omega}$. Then there exists a* smallest *σ-algebra $\sigma(\mathcal{E})$ with $\mathcal{E} \subset \sigma(\mathcal{E})$:*

$$\sigma(\mathcal{E}) := \bigcap_{\substack{\mathcal{A} \subset 2^{\Omega} \text{ is a } \sigma\text{-algebra} \\ \mathcal{E} \subset \mathcal{A}}} \mathcal{A}.$$

$\sigma(\mathcal{E})$ is called the σ-algebra generated by $\mathcal{E}$. $\mathcal{E}$ is called a generator of $\sigma(\mathcal{E})$. Similarly, we define $\delta(\mathcal{E})$ as the λ-system generated by $\mathcal{E}$.

Theorem 2.3 ($\cap$-closed λ-system) *Let $\mathcal{D} \subset 2^{\Omega}$ be a λ-system. Then*

$$\mathcal{D} \text{ is a } \pi\text{-system} \iff \mathcal{D} \text{ is a } \sigma\text{-algebra}$$

Theorem 2.4 (Dynkin's π-λ theorem) *If $\mathcal{E} \subset 2^{\Omega}$ is a π-system, then*

$$\sigma(\mathcal{E}) = \delta(\mathcal{E})$$

Definition 2.7 (liminf and limsup) *Let $A_1, A_2, \ldots$ be subsets of Ω. The sets*

$$\liminf_{n\to\infty} A_n := \bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} A_m \quad \text{and} \quad \limsup_{n\to\infty} A_n := \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m$$

*are called **limes inferior** and **limes superior**, respectively, of the sequence $(A_n)_{n\in\mathbb{N}}$.*

Definition 2.8 (Topology) *Let $\Omega \neq \emptyset$ be an arbitrary set. A class of sets $\tau \subset \Omega$ is called a **topology** on Ω if it has the following three properties:*

- $\emptyset, \Omega \in \tau$
- *τ is $\cap$-closed*
- *$\bigcup_{A\in\mathcal{F}} A \in \tau$ for any $\mathcal{F} \subset \tau$.*

The pair (Ω, τ) is called a topological space. The sets $A \in \tau$ are called open, and the sets $A \subset \Omega$ with $A^c \in \tau$ are called closed.

Definition 2.9 (Borel σ-algebra) *Let (Ω, τ) be a topological space. The σ-algebra*

$$\mathcal{B}(\Omega) := \mathcal{B}(\Omega, \tau) := \sigma(\tau)$$

*that is generated by the open sets is called the **Borel σ-algebra** on Ω. The elements $A \in \mathcal{B}(\Omega, \tau)$ are called **Borel sets** or **Borel measurable sets**.*

Definition 2.10 *Let $\mathcal{A} \subset 2^{\Omega}$ and let $\mu : \mathcal{A} \to [0, \infty]$ be a set function. We say that μ is*

1. ***monotone** if $\mu(A) \leq \mu(B)$ for any two sets $A, B \in \mathcal{A}$ with $A \subset B$,*
2. ***additive** if $\mu(\bigcup\limits_{i=1}^{n})A_i = \sum\limits_{i=1}^{n} \mu(A_i)$ for any choice of* finitely *many mutually disjoint sets $A_1, \ldots, A_n \in \mathcal{A}$ with $A_i \in \mathcal{A}$,*
3. ***σ-additive** if $\mu(\bigcup\limits_{i=1}^{\infty} A_i) = \sum\limits_{i=1}^{\infty} \mu(A_i)$ for any choice of* countably *many mutually disjoint sets $A_1, A_2, \cdots \in \mathcal{A}$ with $A_i \in \mathcal{A}$,*
4. ***subadditive** if for any choice of* finitely *many sets $A, A_1, \ldots, A_n \in \mathcal{A}$ with $A \subset \bigcup\limits_{i=1}^{n} A_i$, we have $\mu(A) \leq \sum\limits_{i=1}^{n} A_i$, and*

5. ***σ-subadditive*** *if for any choice of* countably *many sets* $A, A_1, A_2, \cdots \in \mathcal{A}$ *with* $A \subset \bigcup_{i=1}^{\infty} A_i$*, we have* $\mu(A) \leq \sum_{i=1}^{\infty} \mu(A_i)$.

Definition 2.11 *Let $\mathcal{A}$ be a semiring (as definition 2.4 indicates) and let $\mu : \mathcal{A} \to [0, \infty]$ be a set function with $\mu(\emptyset) = 0$. μ is called*

1. ***content*** *if μ is additive,*
2. ***premeasure*** *if μ is σ-additive,*
3. ***measure*** *if μ is a premeasure and $\mathcal{A}$ is a σ-algebra, and*
4. ***probability measure*** *if μ is a measure and $\mu(\Omega) = 1$.*

Definition 2.12 *Let $\mathcal{A}$ be a semiring. A content μ on $\mathcal{A}$ is called*

1. ***finite*** *if $\mu(A) < \infty$ for every $A \in \mathcal{A}$ and*
2. ***σ-finite*** *if there exists a sequence of sets $\Omega_1, \Omega_2, \cdots \in \mathcal{A}$ such that $\Omega = \bigcup_{n=1}^{\infty} \Omega_n$ and $\mu(\Omega_n) < \infty$ for all $n \in \mathbb{N}$.*

Example 2.1 (Contents and Measure) In this example part, we will introduce some common conceptions that could be met with in the future learning.

1. Let Ω be an (at most) countable nonempty set and let $\mathcal{A} = 2^{\Omega}$. Further, let $(p_\omega)_{\omega \in \Omega}$ be nonnegative numbers. Then $A \mapsto \mu(A) := \sum_{\omega \in A} p_\omega$ defines a σ-finite measure on 2^{Ω}. We call $p = (p_\omega)_{\omega \in \Omega}$ the **weight function** of μ. The number p_ω is called the weight of μ at point ω.
2. If $\sum_{\omega \in \Omega} p_\omega = 1$, then μ is a probability measure. In this case, we interpret p_ω as the probability of the elementary event ω. The vector $p = (p_\omega)_{\omega \in \Omega}$ is called a **probability vector**.
3. Let Ω be a finite nonempty set. By
$$\mu(A) = \frac{\#A}{\#\Omega} \quad \text{for } A \subset \Omega$$
we define a probability measure on $\mathcal{A} = 2^{\Omega}$. This μ is called the uniform distribution on Ω. For this distribution, we introduce the symbol $\mathcal{U}_\Omega := \mu$. The resulting triple $(\Omega, \mathcal{A}, \mathcal{U}_\Omega)$ is called a **Laplace space**.
4. Let $f : \mathbb{R} \to [0, \infty)$ be continuous. In a similar way, we define
$$\mu_f(A) = \sum_{i=1}^{n} \int_{a_i}^{b_i} f(x)dx.$$
Then μ_f is a σ-finite content on $\mathcal{A}$ (even a premeasure). The function f is called the **density** of μ and plays a role similar to the weight function p.

Lemma 2.1 (Properties of contents) *Let $\mathcal{A}$ be a semiring and let μ be a content on $\mathcal{A}$. Then the following statements hold.*

1. *If $\mathcal{A}$ is a ring, then $\mu(A \cup B) + \mu(A \cap B) = \mu(A) + \mu(B)$ for any two sets $A, B \in \mathcal{A}$.*
2. *μ is monotone. If $\mathcal{A}$ is a ring, then $\mu(B) = \mu(A) + \mu(B \backslash A)$ for any two sets $A, B \in \mathcal{A}$ with $A \subset B$.*
3. *μ is subadditive. If μ is σ-additive, then μ is also σ-additive.*
4. *If $\mathcal{A}$ is a ring, then $\sum_{n=1}^{\infty} \mu(A_n) \leq \mu(\bigcup_{n=1}^{\infty} A_n)$ for any choice of countably many mutually disjoint sets $A_1, A_2, \cdots \in \mathcal{A}$ with $\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$.*

Theorem 2.5 (Inclusion-exclusion formula) *Let $\mathcal{A}$ be a ring and let μ be a content on $\mathcal{A}$. Let $n \in \mathbb{N}$ and $A_1, \ldots, A_n \in \mathcal{A}$. Then the following inclusion and exclusion formulas hold:*

$$\mu(A_1 \cup \cdots \cup A_n) = \sum_{k=1}^{n} (-1)^{k-1} \sum_{\{i_1,\ldots,i_k\} \subset \{1,\ldots,n\}} \mu(A_1 \cap \cdots \cap A_n)$$

$$\mu(A_1 \cap \cdots \cap A_n) = \sum_{k=1}^{n} (-1)^{k-1} \sum_{\{i_1,\ldots,i_k\} \subset \{1,\ldots,n\}} \mu(A_1 \cup \cdots \cup A_n)$$

Definition 2.13 *Let $A, A_1, A_2, \ldots$ be sets. We write*

- *$A_n \uparrow A$ and say that $(A_n)n \in \mathbb{N}$ increases to A if $A_1 \subset A_2 \subset \ldots$ and $\bigcup_{n=1}^{\infty} A_n = A$, and*
- *$A_n \downarrow A$ and say that $(A_n)n \in \mathbb{N}$ decreases to A if $A_1 \supset A_2 \supset \ldots$ and $\bigcup_{n=1}^{\infty} A_n = A$.*

Definition 2.14 (Continuity of contents) *Let μ be a content on the ring $\mathcal{A}$.*

1. *μ is called **lower semicontinuous** if $\lim_{n\to\infty} \mu(A_n) = \mu(A)$ for any $A \in \mathcal{A}$ and any sequence $(A_n)_{n\in\mathbb{N}}$ in $\mathcal{A}$ with $A_n \uparrow A$.*
2. *μ is called **upper semicontinuous** if $\lim_{n\to\infty} \mu(A_n) = \mu(A)$ for any $A \in \mathcal{A}$ and any sequence $(A_n)_{n\in\mathbb{N}}$ in $\mathcal{A}$ with $\mu(A_n) < \infty$ for some (and then eventually for all) $n \in \mathbb{N}$ and $A_n \downarrow A$.*
3. *μ is called **$\emptyset$-continuous** if (2) holds for $A = \emptyset$.*

Theorem 2.6 (Continuity and premeasure) *Let μ be a content on the ring $\mathcal{A}$. Consider the following five properties.*

1. *μ is σ-additive (and hence a premeasure).*
2. *μ is σ-subadditive.*
3. *μ is lower semicontinuous.*
4. *μ is $\emptyset$-continuous.*
5. *μ is upper semicontinuous.*

Then the following implications hold:

$$(1) \iff (2) \iff (3) \implies (4) \iff (5)$$

if μ is finite*, then we also have $(4) \implies (3)$.*

Definition 2.15 *We provide the definition of probability space and events.*

1. *A pair $(\Omega, \mathcal{A})$ consisting of a nonempty set Ω and a σ-*algebra *$\mathcal{A} \subset 2^{\Omega}$ is called a **measurable space** (NOT measure space). The sets $A \in \mathcal{A}$ are called **measurable sets**. If Ω is at most countably infinite and if $\mathcal{A} = 2^{\Omega}$, then the measurable space $(\Omega, 2^{\Omega})$ is called **discrete**.*
2. *A triple $(\Omega, \mathcal{A}, \mu)$ is called a **measure space** if $(\Omega, \mathcal{A})$ is a measurable space and if μ is a measure on $\mathcal{A}$.*
3. *If in addition $\mu(\Omega) = 1$, then $(\Omega, \mathcal{A}, \mu)$ is called a **probability space**. In this case, the sets $A \in \mathcal{A}$ are called **events**.*
4. *The set of all finite measures on $(\Omega, \mathcal{A})$ is denoted by $\mathcal{M}_f(\Omega) := \mathcal{M}_f(\Omega, \mathcal{A})$. The subset of probability measures is denoted by $\mathcal{M}_1(\Omega) := \mathcal{M}_1(\Omega, \mathcal{A})$. Finally, the set of σ-finite measures on $(\Omega, \mathcal{A})$ is denoted by $\mathcal{M}_\sigma(\Omega, \mathcal{A})$.*

2.2 Measure Extension Theorem

The main result of this subsection is Caratheodory's measure extension theorem.

Theorem 2.7 (Caratheodory) *Let $\mathcal{A} \subset 2^\Omega$ be a ring and let μ be a σ-finite premeasure on $\mathcal{A}$. There exists a unique measure $\tilde{\mu}$ on $\sigma(\mathcal{A})$ such that $\tilde{\mu}(A) = \mu(A)$ for all $A \in \mathcal{A}$. Furthermore, $\tilde{\mu}$ is σ-finite.*

We prepare for the proof of this theorem with a couple of lemmas.

Definition 2.16 (Uniqueness by an $\cap$-closed generator) *Let $(\Omega, \mathcal{A}, \mu)$ be a σ-finite measure space and let $E \subset \mathcal{A}$ be a π-system that generates $\mathcal{A}$. Assume that there exist sets $E_1, E_2, \dots \in \mathcal{E}$ such that $E_n \uparrow \Omega$ and $\mu(E_n) < \infty$ for all $n \in \mathbb{N}$. Then μ is uniquely determined by the values $\mu(E)$, $E \in \mathcal{E}$. (If μ is a probability measure, the existence of the sequence $(E_n)_{n\in\mathbb{N}}$ is not needed)*

Example 2.2 (Distribution function) A probability measure μ on the space $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ is uniquely determined by the values $\mu((-\infty, b])$ (where $(-\infty, b] = \bigotimes_{i=1}^{n}(-\infty, b_i], b \in \mathbb{R}_n$). In fact, these sets form a π-system that generates $\mathcal{B}(\mathbb{R}_n)$. In particular, a probability measure μ on $\mathbb{R}$ is uniquely determined by its **distribution function** $F : \mathbb{R} \to [0,1], x \mapsto \mu((-\infty, x])$.

Definition 2.17 (Outer measure) *A set function $\mu^* : 2^\Omega \to [0, \infty]$ is called an **outer measure** if*

1. $\mu^*(\emptyset) = 0$,
2. μ^* *is monotone,*
3. μ^* *is σ-additive.*

Lemma 2.2 *Let $\mathcal{A} \subset 2^\Omega$ be an arbitrary class of sets with $\emptyset \in \mathcal{A}$ and let μ be a monotone set function on $\mathcal{A}$ with $\mu(\emptyset) = 0$. For $A \subset \Omega$, define the set of countable coverings of $\mathcal{F}$ with sets $F \in \mathcal{A}$:*

$$\mathcal{U}(A) = \left\{ \mathcal{F} \subset \mathcal{A} : \mathcal{F} \text{ is at most countable and } A \subset \bigcup_{F \in \mathcal{F}} F \right\}.$$

Define

$$\mu^*(A) := \inf \left\{ \sum_{F \in \mathcal{F}} \mu(F) : \mathcal{F} \in \mathcal{U}(A) \right\}.$$

*Then μ^* is an **outer measure**. If in addition μ is σ-subadditive, then $\mu^*(A) = \mu(A)$ for all $A \in \mathcal{A}$.*

Definition 2.18 (μ^* measurable sets) *Let μ^* be an outer measure. A set $A \in 2^\Omega$ is called μ^***-measurable** if*

$$\mu^*(A \cap E) + \mu^*(A^c \cap E) = \mu^*(E) \quad \forall E \in 2^\Omega.$$

We write $\mathcal{M}(\mu^*) = \{A \in 2^\Omega : A \text{ is } \mu^*\text{-measurable}\}$.

Lemma 2.3 *$A \in \mathcal{M}(\mu^*)$ if and only if*

$$\mu^*(A \cap E) + \mu^*(A^c \cap E) \le \mu^*(E) \quad \forall E \in 2^\Omega.$$

Lemma 2.4 *$\mathcal{M}(\mu^*)$ is an algebra.*

Lemma 2.5 *An outer measure μ^* is σ-additive on $\mathcal{M}(\mu^*)$.*

Lemma 2.6 *If μ^* is an outer measure, then $\mathcal{M}(\mu^*)$ is a σ-algebra. In particular, μ^* is a measure on $\mathcal{M}(\mu^*)$.*

Theorem 2.8 (Extension theorem for measures) *Let $\mathcal{A}$ be a semiring and let $\mu : \mathcal{A} \to [0, \infty]$ be an additive, σ-subadditive and σ-finite set function with $\mu(\emptyset) = 0$. Then there is a unique σ-finite measure $\tilde{\mu} : \sigma(\mathcal{A}) \to [0, \infty]$ such that $\tilde{\mu}(A) = \mu(A)$ for all $A \in \mathcal{A}$.*

References

[1] A. KLENKE, *Probability theory: a comprehensive course*, Springer Science & Business Media, 2013.