

Notes of Probability and Stochastics

Xie Zejian

Zhang Songxin

2021-02-04

Contents

0.1	Notations	4
1	Measure and integrations	5
1.1	Measurable space	5
1.2	Measurable function	7
1.3	Measure	12
1.4	Integration	17
1.5	Transforms and Indefinite integral	22
1.6	Kernels and Product spaces	24
2	Probability Spaces	31
2.1	Probability Spaces and Random Variables	31
2.2	Expectation	33

0.1 Notations

\mathbb{R}	$(-\infty, \infty)$
$\overline{\mathbb{R}}$	$[-\infty, \infty]$
\mathbb{R}_+	$[0, \infty)$
\overline{A}	Closure of set A
A°	Interior of set A
$(x_n) \subset A$	A sequence taking value in A
2^A	The power set of A
\mathcal{A}	A collection of subsets in A , i.e., $\mathcal{A} \subset 2^A$
$\ker \mathcal{A}$	$\bigcap_{A \in \mathcal{A}} A$
$x_n \nearrow x$	(x_n) is increasing and converges to x .
$\sigma(\mathcal{A})$	σ -algebra generated by \mathcal{A} .
\mathcal{A}_+	Nonnegative function in \mathcal{A}
$\mu \ll \nu$	μ is absolutely continuous w.r.t. ν .
$\mu f = \int f d\mu = \int f(x) \mu(dx)$	integral
$f: X \rightarrow Y$	x is a function from X to Y .
$f = x \mapsto 5x$	$f(x) = 5x$
$f: X \hookrightarrow Y$	f is an embedding from X to Y .
s.t.	such that
w.r.t.	with respect to
r.v.	random variable

Chapter 1

Measure and integrations

1.1 Measurable space

1.1.1 σ algebra

Definition 1.1. A nonempty system of subset of Ω is an algebra on Ω if it's

1. Closed under complement: $A^c \in \mathcal{A} \iff A \in \mathcal{A}$
2. Closed under finite union: $\cup_i A_i \in \mathcal{A} \iff \forall i, A_i \in \mathcal{A}$

It's an σ algebra on Ω if it's also closed under countable union.

Remark. \mathcal{A} is an algebra auto implies $\emptyset \in \mathcal{A}$ and $\Omega \in \mathcal{A}$. So $\{\emptyset, \Omega\}$ is the minimum algebra on Ω and thus minimum σ algebra while the discrete algebra 2^Ω is maximum.

Let $\{\mathcal{A}_\gamma : \gamma \in \Gamma\}$ is a collection of σ algebra, then we have $\mathcal{A} = \cap_{\gamma \in \Gamma} \mathcal{A}_\gamma$ is also a σ algebra. Hence

Definition 1.2. The smallest σ algebra as intersection of all σ algebras contains \mathcal{A} , that called the σ algebra **generated** by \mathcal{A} and denoted by $\sigma(\mathcal{A})$.

The smallest σ -algebra generated by the system of all open sets in a topological space (Ω, τ) is called **Borel σ algebra** on Ω and denoted by $\mathcal{B}(\Omega)$, its elements are called **Borel sets**.

1.1.2 π, λ, m systems

Definition 1.3. A collection of subsets \mathcal{A} is called.

- **m-system** if closed under monotone series, that is if $(A_n) \subset \mathcal{A}$ and $A_n \nearrow A$, then $A \in \mathcal{A}$.
- **π -system** is closed under finite intersection
- **λ -system** if
 1. $\Omega \in \mathcal{A}$
 2. closed under proper difference:

$$A - B \in \mathcal{A} \iff B \subset A \text{ and } B, A \in \mathcal{A}$$

3. is an m-system.

Theorem 1.1. *Let \mathcal{A} be a collection of subsets of Ω iff it's both a π system and λ system.*

Proof. For \Rightarrow , check:

1. $\Omega \in \mathcal{A}$
2. $A - B = A \cap B^c \in \mathcal{A}$
3. is an m-system

For the converse:

1. $A^c = \Omega - A \in \mathcal{A}$
2. $A \cup B = (A^c \cap B^c)^c \in \mathcal{A}$
3. hence \mathcal{A} is an algebra and \mathcal{A} is a m-system.

Similarly, for m, π, λ -system, they also has a minimum system generated by some collection \mathcal{C} .

□

Lemma 1.1. *Let \mathcal{A} be an algebra, then*

1. $m(\mathcal{A}) = \sigma(\mathcal{A})$
2. if \mathcal{B} is an m class and $\mathcal{A} \subset \mathcal{B}$, then $\sigma(\mathcal{A}) = m(\mathcal{A}) \subset \mathcal{B}$

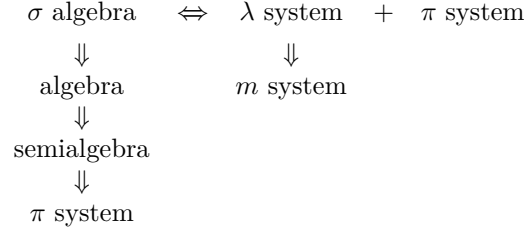
Similarly, let \mathcal{A} be a π class, then $\lambda(\mathcal{A}) = \sigma(\mathcal{A})$

Then we have **Monotone class theorem**:

Theorem 1.2. $\forall \mathcal{A} \subset \mathcal{B} \subset \mathcal{P}(\Omega)$, s.t.:

1. If \mathcal{A} is a π -class, \mathcal{B} is a λ -class, then $\sigma(\mathcal{A}) \subset \mathcal{B}$
2. If \mathcal{A} is an algebra, \mathcal{B} is a m -class, then $\sigma(\mathcal{A}) \subset \mathcal{B}$

1.1.3 Graphical illustration of different classes



1.1.4 Measurable spaces

Definition 1.4 (Measurable Space). Pair (Ω, \mathcal{A}) where \mathcal{A} is a σ -Algebra on Ω .

Definition 1.5 (Products of measurable spaces). Let $(E, \mathcal{E}), (F, \mathcal{F})$ be two measurable spaces. For $A \subset E, B \subset F$, $A \times B$ is the set of all pairs $(x, y) : x \in A, y \in B$. Note that $\mathcal{E} \times \mathcal{F}$ is also a σ -Algebra with all $A \times B$ where $A \in \mathcal{E}, B \in \mathcal{F}$ which is called *the product σ -Algebra*.

1.2 Measurable function

1.2.1 Mappings

Let $f : \Omega_1 \rightarrow \Omega_2$ be a mapping, $\forall B \subset \Omega_2$ and $\mathcal{G} \subset \mathcal{P}(\Omega_2)$, the **inverse image** of

- B is $f^{-1}(B) = \{\omega : \omega \in \Omega_1, f(\omega) \in B\} := \{f \in B\}$
- \mathcal{G} is $f^{-1}(\mathcal{G}) = \{f^{-1}(B) : B \in \mathcal{G}\}$

There is some properties:

1. $f^{-1}(\Omega_2) = \Omega_1, f^{-1}(\emptyset) = \emptyset$
2. $f^{-1}(B^c) = [f^{-1}(B)]^c$
3.
$$\begin{aligned}
 f^{-1}\left(\bigcup_{\gamma \in \Gamma} B_\gamma\right) &= \bigcup_{\gamma \in \Gamma} f^{-1}(B_\gamma) \text{ for } B_\gamma \subset \Omega_2, \gamma \in \Gamma \\
 f^{-1}\left(\bigcap_{\gamma \in \Gamma} B_\gamma\right) &= \bigcap_{\gamma \in \Gamma} f^{-1}(B_\gamma) \text{ for } B_\gamma \subset \Omega_2, \gamma \in \Gamma
 \end{aligned}$$

where Γ may not countable.

4. $f^{-1}(B_1 - B_2) = f^{-1}(B_1) - f^{-1}(B_2) \forall B_1, B_2 \subset \Omega_2$
5. $B_1 \subset B_2 \subset \Omega_2 \implies f^{-1}(B_1) \subset f^{-1}(B_2)$

6. If \mathcal{B} is a σ -algebra, $f^{-1}(\mathcal{B})$ is also a σ -algebra. It's easy to check $f^{-1}(\mathcal{B})$ is closed under complement and countable union. (From properties 2 and 3)
7. If \mathcal{C} is nonempty, $f^{-1}(\sigma(\mathcal{C})) = \sigma(f^{-1}(\mathcal{C}))$

Remarks f^{-1} preserves all the set operations on Ω .

1.2.2 Measurable functions

Definition 1.6. For two measurable spaces (Ω_1, \mathcal{A}) , (Ω_2, \mathcal{B}) , $f : \Omega_1 \rightarrow \Omega_2$ is a **measurable mapping** if $f^{-1}(\mathcal{B}) \subset \mathcal{A}$, where

$$f^{-1}(\mathcal{B}) = \{f^{-1}(B) : B \in \mathcal{B}\}$$

It is a **measurable function** if $(\Omega_2, \mathcal{B}) = (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$, moreover, a **Borel function** if $(\Omega_1, \mathcal{A}) = (\mathbb{R}^m, \mathcal{B}(\mathbb{R}^m))$

Remark. If $\mathcal{B} = \sigma(\mathcal{C})$, the definition can be reduced to $f^{-1}(\mathcal{C}) \subset \mathcal{A}$ since

$$f^{-1}(\mathcal{B}) = f^{-1}(\sigma(\mathcal{C})) = \sigma(f^{-1}(\mathcal{C})) \subset \sigma(\mathcal{A}) = \mathcal{A}$$

Lemma 1.2. Suppose $f : \mathcal{E} \rightarrow \mathcal{F}$ and $g : \mathcal{F} \rightarrow \mathcal{G}$ are measurable, then so is $f \circ g$.

Proof. The same as how we proved composition of continuous function is continuous. □

1.2.3 Random Variable

A r.v. X is a measurable function from (Ω_1, \mathcal{A}) to $(\mathbb{R}, \mathcal{B})$. It denoted by X is \mathcal{A} -measurable or $X \in \mathcal{A}$

(Another definition): A r.v. X is a measurable mapping from (Ω, \mathcal{A}, P) to $(\mathbb{R}, \mathcal{B})$ such that

$$P(|X| = \infty) = P(\{\omega : |X(\omega)| = \infty\}) = 0$$

Lemma 1.3. X is a r.v. from (Ω, \mathcal{A}) to $(\mathbb{R}, \mathcal{B})$

$$\Leftrightarrow X \leq x = X^{-1}([-\infty], x) \in \mathcal{A} \quad \forall x \in \mathbb{R}$$

$$\Leftrightarrow X \leq x = X^{-1}([-\infty], x) \in \mathcal{A} \quad \forall x \in D$$

where D is a dense subset of \mathbb{R} , e.g. \mathbb{Q} . $\{X \leq x\}$ above can be replaced by

$$\{X \leq x\}, \quad \{X \geq x\}, \quad \{X < x\}, \quad \{X > x\}, \quad \{x < X < y\}$$

1.2.4 Construction of random variables

Lemma 1.4. $\mathbf{X} = (X_1, \dots, X_n)$ is a random vectors if X_k is a r.v. $\forall k$ iff \mathbf{X} is a measurable function from (Ω, \mathcal{A}) to $(\mathcal{R}^n, \mathcal{B}(\mathcal{R}^n))$.

Proof. Note that

$$\{\mathbf{X} \in \prod I_n\} = \bigcap \{X_n \in I_n\} \in \mathcal{A}$$

where $I_k = (a_k, b_k]$, $-\infty \leq a_k \leq b_k \leq \infty$ and follows from $\sigma(\{\prod I_n\}) = \mathcal{B}(\mathcal{R}^n)$. For the other direction, note

$$\{X_k \leq t\} = \{\mathbf{X} \in \prod_{i < k} \mathbb{R} \times \{-\infty, t\} \times \prod_{i > k} \mathbb{R}\} \in \mathcal{A}$$

□

Recall lemma 1.2 we have:

Theorem 1.3. \forall random n vectors $X = (X_{1:n})$ and Borel function f from $\mathcal{R}^n \rightarrow \mathcal{R}^m$, then $f(X)$ is a random m vectors.

Remark. Note that continuous function are borel measurable since continuity leads to inverse image of an open set is still open. So if $X_{1:n}$ are r.v.'s, so are $\sum X_n$, $\sin(x)$, e^X , $\text{Poly}(X)$, \dots . That implies:

$\forall X, Y \in \mathcal{A}$, so are $aX + bY$, $X \vee Y = \max\{X, Y\}$, $X \wedge Y = \min\{X, Y\}$, X^2 , XY , X/Y , $X^+ = \max(x, 0)$, $X^- = -\min(x, 0)$, $|X| = X^+ + X^-$

1.2.5 Limiting opts

Let (X_n) are r.v. on (Ω, \mathcal{A}) , then $\sup_{n \rightarrow \infty} X_n$, $\inf_{n \rightarrow \infty} X_n$, $\limsup_{n \rightarrow \infty} X_n$, $\liminf_{n \rightarrow \infty} X_n$ are r.v.'s. Moreover, if it exists, $\lim_{n \rightarrow \infty} X_n$ is r.v..

Proof. First two follows from, $\forall t \in \mathbb{R}$:

$$\begin{aligned} \left\{ \sup_{n \rightarrow \infty} X_n \leq t \right\} &= \bigcap_{n=1}^{\infty} \{X_n \leq t\} \in \mathcal{A} \\ \left\{ \inf_{n \rightarrow \infty} X_n \geq t \right\} &= \bigcap_{n=1}^{\infty} \{X_n \geq t\} \in \mathcal{A} \end{aligned}$$

and the last two follows from $\limsup_{n \rightarrow \infty} = \inf_{k \rightarrow \infty} \sup_{m \geq k} X_m$ and $\liminf_{n \rightarrow \infty} = \sup_{k \rightarrow \infty} \inf_{m \geq k} X_m$.

□

That implies

Lemma 1.5. *If $S = \sum_1^\infty X_n$ exists everywhere, then S is a r.v.*

Proof. Note $\sum_1^\infty X = \lim_{n \rightarrow \infty} \sum_n X_n$ is a r.v.

□

If $X = \lim_{n \rightarrow \infty} X_n$ holds **almost** everywhere, i.e., define Ω_0 is the set of all ω , such that $\lim_n X_n(\omega)$ exists, then $P(\Omega_0) = 1$, we say that X_n converges a.s. and write:

$$X_n \rightarrow X \quad a.s.$$

For a measurable function f , we may modify it at a null set into f' and it remain measurable since for any open set G , $f'^{-1}(G)$ differ $f^{-1}(G)$ at most null set, by the completion of Lebesgue measure space, $f'^{-1}(G)$ is measurable and thus f'^{-1} measurable. Hence, for $f_n \rightarrow f$ a.s., we may ignore a null set and then $f_n \rightarrow f$ everywhere and thus f measurable.

1.2.6 Approximations of r.v. by simple r.v.'s

Definition 1.7. If $A \in \mathcal{A}$, the indicator function $\mathbf{1}_A$ is a r.v. If $\Omega = \sum_1^n A_i$, where $A_i \in \mathcal{A}$, then $X = \sum_1^n a_i \mathbf{1}_{A_i}$ is a r.v. and called **simple r.v.**

Any r.v. can be approximated by simple ones:

Theorem 1.4. $\forall X \in \mathcal{A}$, $\exists 0 \leq X_1 \leq X_2 \leq \dots \leq X_n$ s.t. $X_n(\omega) \nearrow X(\omega)$ everywhere.

Proof. Suppose

$$X_n(\omega) = \sup\left\{\frac{j}{2^n} : j \in \mathbb{Z}, \frac{j}{2^n} \leq \min(X(\omega), 2^n)\right\}$$

One can check X_n is simple r.v. and $X_n(\omega) \nearrow X(\omega)$ for all $\omega \in \Omega$.

□

1.2.7 σ algebra generated by r.v.

Let $\{X_\lambda, \lambda \in \Lambda\}$ is r.v.s on (Ω, \mathcal{A}) . Define

$$\sigma(X_\lambda, \lambda \in \Lambda) := \sigma(X_\lambda \in B, B \in \mathcal{B}, \lambda \in \Lambda) = \sigma(X_\lambda^{-1}(\mathcal{B}), \lambda \in \Lambda) = \sigma(\cup_{\lambda \in \Lambda} X_\lambda^{-1}(\mathcal{B}))$$

which is called σ algebra generated by $\{X_\lambda, \lambda \in \Lambda\}$, where Λ is a index set which can be uncountable.

For $\Lambda = \mathbb{N}^+$:

1.

$$\begin{aligned}\sigma(X_i) &= \sigma(X_i^{-1}(\mathcal{B})) = X_i^{-1}(\mathcal{B}) = \{X_i \in \mathcal{B}\} \\ \sigma(X_1, \dots, X_n) &= \sigma(\cup_{i=1}^n X_i^{-1}(\mathcal{B})) = \sigma(\cup_{i=1}^n \sigma(X_i))\end{aligned}$$
2.

$$\begin{aligned}\sigma(X_1) &\subset \sigma(X_1, X_2) \subset \dots \subset \sigma(X_1, \dots, X_n) \\ \sigma(X_1, X_2, \dots) &\supset \sigma(X_2, X_3, \dots) \supset \dots \supset \sigma(X_n, X_{n+1}, \dots)\end{aligned}$$
3. $\bigcap_1^\infty \sigma(X_n, X_{n+1}, \dots)$ is the tail σ algebra of X_1 .

If $A_{1:n}$ are not mutually exclusive to each other, then we have

$$|\sigma(A_{1:n})| = 2^{2^n}$$

Which follows from for a partition $A_{1:n}$,

$$\sigma(A_1, \dots, A_n) = \left\{ \bigcup_{i \in J} A_i \right\}$$

where J is any subset of $\mathbb{N} \leq n$ and $A_0 = \emptyset$. Hence for discrete r.v. Y , $\sigma(Y)$ can be generated from $A_i = \{Y = y_i\}$ for all y_i 's. For continuous case, it's generated by all intervals.

1.2.8 Monotone classes of function

Definition 1.8 (monotone class). \mathcal{M} is called a monotone class if:

- $1 \in \mathcal{M}$
- $f, g \in \mathcal{M}_b$ and $a, b \in \mathbb{R} \implies af + bg \in \mathcal{M}$
- $(f_n) \subset \mathcal{M}_+, f_n \uparrow f \implies f \in \mathcal{M}$

where \mathcal{M}_+ is a subcollection consisting of positive functions in \mathcal{M} , and \mathcal{M}_b is the bounded function in \mathcal{M} .

Theorem 1.5 (Monotone class theorem for functions). *Let \mathcal{M} be a monotone class of functions on (Ω, \mathcal{A}) . Suppose for some π -system \mathcal{C} generating \mathcal{A} and $1_A \in \mathcal{M}$ for every $A \in \mathcal{C}$. Then $\mathcal{A}_+, \mathcal{A}_b \subset \mathcal{M}$*

Proof. First we need to show that $1_A \in \mathcal{M}$ for every $A \in \mathcal{A}$. Let $\mathcal{D} = \{A \in \mathcal{A} : 1_A \in \mathcal{M}\}$. Now we check that \mathcal{D} is a λ -system:

- $1_\Omega = 1$, so $\Omega \in \mathcal{D}$.
- $B \subset A, A, B \in \mathcal{D}$. $1_{A-B} = 1_A - 1_B \in \mathcal{D}$
- $(A_n) \subset \mathcal{D}$ and $A_n \uparrow A$, then $1_{A_n} \uparrow 1_A$, so $1_A \in \mathcal{M}$, then $A \in \mathcal{D}$

By assumption, $\mathcal{C} \subset \mathcal{D}$, and $\sigma(\mathcal{C})$ is the smallest d-system by the theorem above, so $\mathcal{E} \subset \mathcal{D}$, so $1_A \in \mathcal{M}$ for every $A \in \mathcal{E}$.

As $1_A \in \mathcal{M}$ for every $A \in \mathcal{E}$, we can easily prove that all of the positive simple function is generated by the linear combination of 1_A s. And all positive \mathcal{E} -measurable functions is generated by a sequence of positive simple functions. Then for general bounded \mathcal{E} -measurable function f , using $f = f^+ - f^-$ where $f^+, f^- \in \mathcal{M}$.

□

Definition 1.9. Let $(E, \mathcal{E}), (F, \mathcal{F})$ be two measurable spaces and f is a bijection $E \rightarrow F$. Then f is said to be a isomorphism of (E, \mathcal{E}) and (F, \mathcal{F}) if f is \mathcal{E} -measurable and f^{-1} is \mathcal{F} -measurable. These two spaces are called isomorphic if there exists an isomorphisms between them.

Definition 1.10. A measurable space (Ω, \mathcal{A}) is said to be *standard* if it there exist an embedding $f : (\Omega, \mathcal{A}) \hookrightarrow (\mathbb{R}, \mathcal{B})$.

Remark. Clearly, $([0, 1], \mathcal{B}([0, 1]))$, $(\mathbb{N} \leq n, 2^{N \leq n})$ and $(\mathbb{N}, 2^{\mathbb{N}})$ are all standard. In fact, every standard measurable space is isomorphic to one of them.

1.3 Measure

Let Ω be a space and \mathcal{A} a class, then function $\mu : \mathcal{A} \rightarrow R = [-\infty, \infty]$ is a **set function**.

It's

- 1. **finite** if $\forall A \in \mathcal{A}, |\mu(A)| < \infty$
 2. **σ -finite** if $\exists A_n \subset \mathcal{A}, s.t. \cup_{i=1}^{\infty} A_i = \Omega \quad \forall n \quad |\mu(A_n)| < \infty$
 3. **s finite** if there exist countable finite (μ_n) s.t. $\mu = \sum_n \mu_n$.
- 1. **additive** $\iff \mu(\sum_{i=1}^n A_i) = \sum_{i=1}^n \mu(A_i)$
 2. **σ -additive** $\iff \mu(\sum_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mu(A_i)$

Remark. Finite implies σ finite and σ finite implies Σ finite.

μ is a **measure** on \mathcal{A} if

1. $\forall A \in \mathcal{A} : \mu(A) \geq 0$
2. It's σ additive.

the triplet $(\Omega, \mathcal{A}, \mu)$ is a **measure space** when μ is a measure and (Ω, \mathcal{A}) is a measurable space. Whose sets are called **measurable sets** or **\mathcal{A} -measurable**. A measure space is a **probability space** if $P(\Omega) = 1$.

Assume that $A_{1:n} \in \mathcal{A}$ and $A \in \mathcal{A}$ and μ is a measure.

1. μ is continues from above, if $A_n \searrow A \implies \mu(A_n) \rightarrow \mu(A)$
2. μ is continues from below, if $A_n \nearrow A \implies \mu(A_n) \rightarrow \mu(A)$
3. μ is continues at A , if $A_n \rightarrow A \implies \mu(A_n) \rightarrow \mu(A)$

\forall Measure μ is continues from below and may not continues from above. It will be continues from above if $\exists m < \infty, \mu(A_m) < \infty$. So finite measure μ are always continues.

1.3.1 Properties of measure

1.3.1.1 Semialgebras

Let μ be a nonnegative additive set function on a semialgebra \mathcal{A} . $\forall A, B \in \mathcal{A}$ and $\{A_n, B_n, n \geq 1\} \in \mathcal{A}$

1. **(Monotonicity):** $A \subset B \implies \mu(A) \leq \mu(B)$
2. **(σ -subadditivity):**
 1. $\sum_{n=1}^{\infty} A_n \subset A, \implies \sum_{n=1}^{\infty} \mu(A_n) \leq \mu(A)$
 2. Moreover, if μ is a measure, then

$$B \subset \sum_{n=1}^{\infty} B_n \implies \mu(B) \leq \sum_{n=1}^{\infty} \mu(B_n)$$

We can assert a nonnegative set function μ is a measure by:

1. μ is additive
2. μ is σ subadditive on \mathcal{S}

1.3.1.2 Algebras

Algebras enjoy all the properties of semialgebra, also, we have

Theorem 1.6 (σ subadditivity). *Let μ be a measure on an algebra \mathcal{A}*

$$A \subset \cup_1^{\infty} A_n \implies \mu(A) \leq \sum_1^{\infty} \mu(A_n)$$

Proof. Note $A = A \cap (\cup A_n) = \cup(A \cap A_n)$, hence we can write A as union in \mathcal{A} by take $B_n = A \cap A_n \in \mathcal{A}$.

$$A = \cup_1^{\infty} B_n$$

and then we can take $C_n = B_n - \cup_1^{n-1} B_i \in \mathcal{A}$ to write A as disjoint union:

$$A = \sum C_n$$

Then

$$\mu(\mathcal{A}) = \mu(\sum C_n) = \sum \mu(C_n) \leq \sum \mu(B_n) \leq \sum \mu(A_n)$$

as $C_n \subset B_n \subset A_n$.

□

1.3.1.3 σ algebras

Let μ be a measure on an σ algebra \mathcal{A}

1. Monotonicity
2. Boole's inequality(Countable Sub-Additivity)

$$\mu(\cup_{i=1}^{\infty} A_i) \leq \sum_{i=1}^{\infty} \mu(A_i)$$

3. Continuity from below
4. Continuity from above if μ is finite in A_i .

The sense of 4 follows from suppose $A_i \searrow A$, then $A_1 - A_i \nearrow A_1 - A$, then

$$\mu(A_1) - \mu(A) = \mu(A_1 - A) = \lim \mu(A_1 - A_i) = \mu(A_1) - \lim(\mu(A_i))$$

where $\mu(A_1)$ cannot be cancelled if $\mu(A_i) = \infty$.

Definition 1.11. Let $(\Omega, \mathcal{A}, \mu)$ be a measure space, and $N \subset \Omega$

1. N is a μ **null set** iff $\exists B \in \mathcal{A}$ s.t. $\mu(B) = 0$, $N \subset B$
2. This measure space is a **complete measure space** if $\forall \mu$ null space N , $N \in \mathcal{A}$

Theorem 1.7. Given any measure space $(\Omega, \mathcal{A}, \mu)$, there exist a complete measure space $(\Omega, \bar{\mathcal{A}}, \bar{\mu})$, such that $\mathcal{A} \subset \bar{\mathcal{A}}$ and $\bar{\mu}$ is an extension of μ . This space is called completion of $(\Omega, \mathcal{A}, \mu)$.

Proof. Take

$$\begin{aligned}\bar{\mathcal{A}} &= \{A \cup N : A \in \mathcal{A}\} \\ \bar{\mathcal{B}} &= \{A \Delta N : A \in \mathcal{A}\}\end{aligned}$$

$\bar{\mathcal{A}} = \bar{\mathcal{B}}$ since $A \cup N = (A - B) \Delta (B \cap (A \cup N))$ and $A \Delta N = (A - B) \cup (B \cap (A \Delta N))$.

Then we can show that $\bar{\mathcal{A}}$ is a σ algebra. Let $\Omega_i = A_i \cup N_i \in \bar{\mathcal{A}}$, then

$$\bigcup_1^\infty \Omega_i = \bigcup_1^\infty A_i \cup \bigcup_1^\infty N_i$$

and note $\bigcup_1^\infty A_i \in \mathcal{A}$ and $\mu(\bigcup_1^\infty N_i) \leq \mu(\bigcup_1^\infty B_i) \leq \bigcup_1^\infty \mu(B_i) = 0$. Thus $\bar{\mathcal{A}}$ is closed by countable union. As for complements, note $\Omega^c = A^c \cap N^c = (A^c \cap N^c \cap B^c) \cup (A^c \cap N^c \cap B) = (A^c \cap B^c) \cup (A^c \cap N^c \cap B) \in \bar{\mathcal{A}}$.

Finally we define a measure $\bar{\mu}$ on $\bar{\mathcal{A}}$ by

$$\bar{\mu}(A \cup N) = \mu(A)$$

We should prove it's well defined. Suppose $A_1 \cup N_1 = A_2 \cup N_2 \in \bar{\mathcal{A}}$, note $A \Delta B \Delta C = A \Delta (B \Delta C)$ and $A \Delta B = B \Delta A$.

$$\begin{aligned}(A_1 \Delta A_2) \Delta (N_1 \Delta N_2) &= (A_1 \Delta A_2 \Delta N_1) \Delta N_2 \\ &= (A_1 \Delta N_1) \Delta (A_2 \Delta N_2) \\ &= \emptyset\end{aligned}$$

Hence $A_1 \Delta A_2 = N_1 \Delta N_2$, note $N_1 \Delta N_2 \subset N_1 \cup N_2 \subset B_1 \cup B_2$, hence $\mu(A_1 \Delta A_2) = 0$ and thus $\mu(A_1 - A_2) = \mu(A_2 - A_1) = 0$. Therefore

$$\begin{aligned}\mu(A_1) &= \mu(A_1 - A_2) + \mu(A_1 \cap A_2) = \mu(A_1 \cap A_2) \\ \mu(A_2) &= \mu(A_2 - A_1) + \mu(A_1 \cap A_2) = \mu(A_1 \cap A_2)\end{aligned}$$

$\bar{\mu}$ is do well defined. μ^* is auto σ additive since so is μ and is easy to check that all μ^* null set is μ null set.

□

1.3.2 Specification of measures

Theorem 1.8. *Let (Ω, \mathcal{A}) be a measurable space and μ, ν be finite measures. If μ, ν agree on a π system generating \mathcal{A} , then μ, ν are identical.*

If μ, ν are just σ finite, then the π system must include the partition $(A_n) \subset \mathcal{A}$.

Proof. Let \mathcal{C} be the π system generating \mathcal{A} and $\mu(A) = \nu(A)$ for every $A \in \mathcal{C}$. Consider $\mathcal{D} = \{A \in \mathcal{A} : \mu(A) = \nu(A)\}$ which satisfies $\mathcal{C} \subset \mathcal{D} \subset \Omega$. Then we need to prove that \mathcal{D} is a λ system:

- $\Omega \in \mathcal{D}$ by the assumption.
- Let $A, B \in \mathcal{D}$ and $B \subset A$. Then $\mu(A-B) = \mu(A) - \mu(B) = \nu(A) - \nu(B) = \nu(A-B)$, so $A-B \in \mathcal{D}$.
- Let $(A_n) \uparrow A$ and $(A_n) \subset \mathcal{D}$, then $\mu(A_n) \uparrow \mu(A)$, $\nu(A_n) \uparrow \nu(A)$, since $\mu(A_n) = \nu(A_n)$ for every n , so $\mu(A) = \nu(A)$.

So \mathcal{D} is a d-system. It follows that $\sigma(\mathcal{C}) = \mathcal{A} \subset \mathcal{D}$.

□

As consequence, we have

Corollary 1.1. *Suppose μ and ν are probability measures on space on $(\mathbb{R}, \mathcal{B})$ then $\mu = \nu$ iff $\mu[-\infty, r] = \nu[-\infty, r], \forall r \in \mathbb{R}$.*

Proof. Note $\{[-\infty, r] : r \in \mathbb{R}\}$ is a π system and generates \mathcal{B} .

□

1.3.3 Atomic and diffuse measure

Definition 1.12. Let $(\Omega, \mathcal{A}, \mu)$ be a measure space where \mathcal{A} contains all the singletons: $\{x\} \in \mathcal{A}$ for every $x \in \Omega$ (it's true for all the standard measure).

A point x is said to be an **atom** if $\mu(\{x\}) > 0$, the measure is said to be **diffuse** if it has no atoms. It is said to be **purely atomic** if the set D of its atoms is countable and $\mu(\Omega - D) = 0$.

Lemma 1.6. *A s-finite measure has at most countable many atoms.*

Proof. It suffices to show that when μ is finite. Suppose $A_n = \{x : \mu\{x\} > \frac{1}{n}\}$ and A consists all atoms, then the claim follows from $A_n \nearrow A$ and $|A_n| \leq n\mu(\Omega)$ as $A = \bigcup_n A_n$.

□

Theorem 1.9. *Let μ be a s-finite measure on (Ω, \mathcal{A}) . Then $\mu = \nu + \lambda$ where λ is a diffuse measure and ν is purely atomic.*

Proof. Let D be set of all atoms and define

$$\begin{aligned}\lambda(A) &= \mu(A - D) \\ \nu(A) &= \mu(A \cap D)\end{aligned}$$

for all $A \in \mathcal{A}$. Clearly, $\lambda + \nu = \mu$. Then

- λ is diffuse as $\lambda\{x\} = 0$ for all $x \in D$ and if $\lambda\{x\} > 0$, it must be $x \in D$.
- ν is purely atomic as $D_\nu = D$ clearly and $\nu(\Omega - D) = \mu(\emptyset) = 0$.

□

sdf

1.4 Integration

let f be Borel measurable on $(\Omega, \mathcal{A}, \mu)$. The **integral** of f w.r.t μ is denoted by

$$\int f(\omega)\mu(d\omega) = \int f d\mu = \int f$$

1. If $f = \sum_1^n a_i \mathbf{1}_{A_i}$ with $a_i \geq 0$,

$$\int f d\mu = \sum_1^n a_i \mu(A_i)$$

2. If $f \geq 0$, define

$$\int f d\mu = \lim_n \int f_n d\mu$$

where f_n are simple functions and $f_n \nearrow f$.

3. For any f , we have $f = f^+ - f^-$, define

$$\int f d\mu := \int f^+ d\mu - \int f^- d\mu$$

4. f is said to be integrable w.r.t. μ if $\int |f| d\mu < \infty$. We denote all integrable functions by L^1 .

Proposition 1.1. (*Integral over sets*)

$$\int_A f d\mu = \int f \mathbf{1}_A d\mu = \int f(\omega) \mathbf{1}_A(\omega) \mu(d\omega)$$

(**Absolute integrability**). $\int f$ is finite iff $\int |f|$ is finite.

(**Linearity**) If $f, g, a, b \geq 0$ or $f, g \in L^1$

$$\int (af + bg) = a \int f + b \int g$$

(σ additivity over sets) If $A = \sum_{i=1}^{\infty} A_i$, then

$$\int_A f = \sum_{i=1}^{\infty} \int_{A_i} f$$

(**Positivity**) If $f \geq 0$ a.s., then $\int f \geq 0$

(**Monotonicity**) If $f_1 \leq f \leq f_2$ a.s., then $\int f_1 \leq \int f \leq \int f_2$

(**Mean value theorem**) If $a \leq f \leq b$ a.s., then

$$a\mu(A) \leq \int_A f \leq b\mu(A)$$

(**Modulus inequality**): $|\int f| \leq \int |f|$

1.4.1 Monotone Convergence Theorem

Theorem 1.10 (Monotone Convergence Theorem). Suppose nonnegative $f_n \nearrow f$ a.e., then $\int f_n d\mu \nearrow \int f d\mu$.

Theorem 1.11. We may ignore a null set then $f_n \nearrow f$ and their integration still equal. Note $\int f_n d\mu \leq \int f d\mu$, $\int f_n d\mu$ must converges to some $L \leq \int f$. Then we show $L \geq \int f$.

Let $s = \sum a_i \chi_{E_i}$ be simple function and $s \leq f$. Let $A_n = \{x : f_n(x) \geq cs(x)\}$ where $c \in (0, 1)$, then $A_n \nearrow X$. For each n

$$\begin{aligned} \int f_n &\geq \int_{A_n} f_n \geq c \int_{A_n} s \\ &= c \int_{A_n} \sum a_i \chi_{E_i} \\ &= c \sum a_i \mu(E_i \cap A_n) \\ &\nearrow c \int s \end{aligned}$$

hence $L \geq c \int s \implies L \geq \int s \implies L = \lim L \geq \lim \int s_n = \int f$.

Lemma 1.7 (Fatou's Lemma). If $f_n \geq 0$ a.e. then

$$\int \left(\liminf_n f_n \right) \leq \liminf_n \int f_n$$

Proof. Suppose $g_n = \inf_{i \geq n} f_i$ and recall that $\lim g_n = \liminf f_n$. Clearly $g_n \leq f_i \forall i \geq n$ hence

$$\int g_n \leq \inf_{i \geq n} \int f_i$$

Take limit both side and note g_n is increasing:

$$\lim \int g_n = \int \lim g_n = \int \liminf f_n \leq \liminf \int f_n$$

□

Theorem 1.12 (Dominated Convergence Theorem). *Suppose $f_n(x) \rightarrow f(x) \forall x$, and there exists a nonnegative integrable g s.t. $|f_n(x)| \leq g(x)$ (then we get $f_n \in L^1$ immediately), then*

$$\lim \int f_n d\mu = \int f d\mu$$

Proof. Since $f_n + g \geq 0$

$$\int f + \int g = \int f + g \leq \liminf \int f_n + g = \liminf \int f_n + \int g$$

thus $\int f \leq \liminf_{n \rightarrow \infty} \int f_n$. Similarly, we can get $\int f \leq \liminf_{n \rightarrow \infty} \int f_n$ from $g - f_n \geq 0$.

□

Theorem 1.13 (Tonelli's Theorem). *If $\sum_1^\infty \int |f_n| < \infty$, then*

$$\int \left(\sum_{i=1}^\infty f_n \right) = \sum_{i=1}^\infty \int f_n$$

Proof. Let $g_k = \sum_1^k |f_n|, g = \sum_1^\infty |f_n|, h_k = \sum_1^k f_n, h = \sum_1^\infty f_n$. Then $g_k \nearrow g$, by MCT, we have

$$\int g = \lim \int g_k = \lim \sum_1^k \int |f_n| = \sum_1^\infty \int |f_n| < \infty$$

Hence we may let g dominate h_k and get

$$\int h = \lim \int h_k = \sum_1^\infty \int f_n$$

□

1.4.2 Criteria for zero a.e.

Lemma 1.8 (Markov inequality). *Let $A = \{x \in \Omega : f(x) \geq M\}$, then*

$$\mu(A) \leq \frac{\int f}{M}$$

Proof.

$$\mu(A) = \int \chi_A = \int_A \chi_A \leq \int_A \frac{f}{M} \leq \int_X \frac{f}{M} = \frac{\int f}{M}$$

□

Lemma 1.9. *Suppose f is measurable and non-negative and $\int f d\mu = 0$. Then $f = 0$ a.e.*

Proof. By lemma 1.8 and define $A_M = \{x \in \Omega : f(x) \geq M\}$. Consequently, $\mu(A_M) = 0$ for all $M > 0$, note $A_{\frac{1}{n}} \nearrow A_0$:

$$A_0 = \bigcup_1^\infty A_{\frac{1}{n}} \implies \mu(A_0) = \sum 0 = 0$$

Hence $f = 0$ a.e.

□

Lemma 1.10. *Suppose f is integrable and $\int_A f = 0$ for all measurable A . Then $f = 0$ a.e.*

Proof. Suppose $A_n = \{x \in \Omega : f(x) \geq \frac{1}{n}\}$, then

$$0 = \int_{A_n} f \geq \frac{\mu(A_n)}{n} \implies \mu(A_n) = 0$$

thus $\mu\{x \in \Omega : f(x) > 0\} = 0$. Similarly, $\mu\{x \in \Omega : f(x) < 0\} = 0$ and the claim follows.

□

Theorem 1.14. *Suppose $f : \mathbb{R} \rightarrow \mathbb{R}$ is integrable and $\int_a^x f = 0$ for all x , then $f = 0$ a.e.*

Proof. For any interval $I = [c, d]$,

$$\int_i f = \int_a^d f - \int_a^c f = 0$$

Then the integral is 0 for finite disjoint union of intervals from additivity. Note open sets G can be written as countable union of disjoint open intervals $G = \sum_1^\infty I_i = \lim \sum I_n \Rightarrow$

$$\int_G f = \int f \chi_G = \int f \sum_1^\infty \chi_{I_i} = \int \lim f \sum \chi_{I_i} = \lim \int f \sum \chi_{I_i} = 0$$

If $G_n \searrow H$, then

$$\int_H f = \int f \chi_H = \int \lim f \chi_{G_n} = \lim \int f \chi_{G_n} = \lim \int_{G_n} f = 0$$

where we apply DMT twice and take dominated function $g = |f|$.

Finally, for any borel measurable set E , there is $G_\delta \supset E$ and $m(G_\delta - E) = 0$, then

$$\int_E f = \int f \chi_E = \int f \chi_{G_\delta} = \int_{G_\delta} f = 0$$

□

1.4.3 Characterization of the integral

Theorem 1.15. *Let (Ω, \mathcal{A}) be a measurable space and $L : \mathcal{A} \rightarrow \overline{\mathbb{R}}_+$, then there is a unique measure μ on (Ω, \mathcal{A}) s.t. $L(f) = \int f$ for every $f \in \mathcal{A}_+$ iff:*

- $f = 0 \Rightarrow L(f) = 0$
- $f, g \in \mathcal{A}_+$ and $a, b \in \mathbb{R}_+ \Rightarrow L(af + bg) = aL(f) + bL(g)$
- $(f_n) \subset \mathcal{A}_+$ and $f_n \nearrow f \Rightarrow L(f_n) \nearrow L(f)$

Proof. \Rightarrow follows from the definition and properties of integral. For \Leftarrow , let there is a function L satisfies above conditions and give a μ and let $\mu(A) = L(1_A)$, then use those conditions we can prove that μ is a measure a (Ω, \mathcal{A}) .

□

1.5 Transforms and Indefinite integral

Definition 1.13 (Image measure). Let (F, \mathcal{F}) and (E, \mathcal{E}) be measurable spaces. Let ν be a measure on (F, \mathcal{F}) and let $h : F \rightarrow E$ be measurable relative to \mathcal{F} and \mathcal{E} , then define a mapping $\nu \circ h^{-1}(B) = \nu(h^{-1}B)$, $B \in \mathcal{E}$. Then $\nu \circ h^{-1}$ is a measure on (E, \mathcal{E}) , which is called the **image** of ν under h .

Remark. Image inherit finite and s-finite, but not σ -finite.

Theorem 1.16. For every $f \in \mathcal{E}$, we have $(\nu \circ h^{-1})f = \nu(f \circ h)$.

Proof. We only need to show that \mathcal{E}_+ case and then the general case follows easily.

Let $L : \mathcal{E}_+ \rightarrow \bar{\mathbb{R}}_+$ by letting $L(f) = \nu(f \circ h)$. Then as the property of $\nu(f \circ h)$, f satisfies the properties of integral characterization theorem. Then, $L(f) = \mu f$ for some unique measure μ on (E, \mathcal{E}) . And note $\mu = \nu \circ h$

$$\mu(B) = L(\mathbf{1}_B) = \nu(\mathbf{1}_B \circ h) = \nu(h^{-1}B)$$

□

1.5.1 Images of the Lebesgue measure

By theorem 1.16, we have a convenient tool for creating new measure from old and, we may integral a measure using Lebesgue measure:

Theorem 1.17. Let $(\Omega, \mathcal{A}, \mu)$ be a standard measure space where μ is s-finite and $b = \mu(\Omega)$. Then there exists a measurable mapping $h : ([0, b], \mathcal{B}_{[0, b]}) \rightarrow (\Omega, \mathcal{A})$ s.t. $\mu = \lambda \circ h^{-1}$, where λ is the Lebesgue measure on $[0, b]$.

Proof. See 5.15 and 5.16 on page 34 in *Probability and Stochastic*.

□

1.5.2 Indefinite integrals

Definition 1.14. Let $(\Omega, \mathcal{A}, \mu)$ be a measure space and $p \in \mathcal{A}_+$. Define:

$$\nu(A) = \mu(p\mathbf{1}_A) = \int_A p d\mu$$

then ν is a measure on (Ω, \mathcal{A}) and called **indefinite integral** of p w.r.t. μ .

Remark. ν is a measure follows from MCT.

Theorem 1.18. For any $f \in \mathcal{A}_+$, $\nu f = \mu(pf)$.

Proof. Let $L(f) = \mu(pf)$. Check L :

- $f = 0 \implies L(f) = 0$
- Give $f, g \in \mathcal{E}_+$ and $a, b \in \mathbb{R}_+ \implies L(af+bg) = \mu(p(af+bg))$ and based on the arithmetic rules on \mathbb{R} and the linearity of μ , $L(af+bg) = aL(f) + bL(g)$
- Give $(f_n) \subset \mathcal{E}_+$ and $f_n \nearrow f$, $L(f_n) = \mu(pf_n)$ and as $f_n \nearrow f$, $pf_n \nearrow pf$ so $\lim L(f_n) = \lim \mu(pf_n)$. According to the monotone converging theorem, $\lim \mu(pf_n) = \mu(pf) = L(f)$

So there exists $\hat{\mu}$ s.t. $\mu(pf) = \hat{\mu}f$ and that force $\hat{\mu} = \nu$ as

$$\hat{\mu}(A) = L(\mathbf{1}_A) = \mu(p\mathbf{1}_A) = \nu(A)$$

□

Remark. Writing above result in an explicit notation:

$$\int_E f d\nu = \int_E pf d\mu$$

that is:

$$d\nu = p d\mu$$

and it's precisely Fundamental theorem of calculus. Thus we can say:

- ν is the indefinite integral of p w.r.t. μ or
- p is the density of ν w.r.t. μ .

1.5.3 Radon-Nikodym theorem

Definition 1.15 (Absolutely continuous). Let ν and μ be measures on a measurable space (Ω, \mathcal{A}) . Then ν is said to be **absolutely continuous** w.r.t. μ if for any set $A \in \mathcal{E}$, $\mu(A) = 0 \implies \nu(A) = 0$ and denoted by $\nu \ll \mu$.

Clearly, if ν is the indefinite integral of some $p \in \mathcal{A}_+$ w.r.t. μ , then it's absolutely continuous w.r.t. μ . And the follows shows that the converse is true.

Theorem 1.19 (Radon-Nikodym Theorem). *Suppose that μ is σ -finite and ν is absolutely continuous w.r.t. μ . Then there exists unique (up to a.e.) $p \in \mathcal{A}_+$ s.t. for every $f \in \mathcal{A}_+$:*

$$\int_{\Omega} f d\nu = \int_{\Omega} fp d\mu$$

1.6 Kernels and Product spaces

Definition 1.16 (transition kernel). Let (E, \mathcal{E}) and (F, \mathcal{F}) be measurable spaces. Let $K : E \times \mathcal{F} \rightarrow \overline{\mathbb{R}}_+$. Then, K is called a transition kernel from (E, \mathcal{E}) into (F, \mathcal{F}) if:

- the mapping $x \mapsto K(x, B)$ is \mathcal{E} -measurable for every set $B \in \mathcal{F}$
- the mapping $B \mapsto K(x, B)$ is a measure on (F, \mathcal{F}) for every $x \in E$

Example 1.1. If ν is a finite measure on (F, \mathcal{F}) , and k is a positive function on $E \times F$ that is $\mathcal{E} \times \mathcal{F}$ -measurable, then

$$K(x, B) = \int_B k(x, y) d\nu$$

when fix $x \in E$, $K(x, B) = \nu(k(x, y)\mathbf{1}_B) = \mu(B)$ for some μ which is the measure on (F, \mathcal{F}) ;

when fix $B \in \mathcal{F}$, $f(x) = K(x, B)$ is measurable follows from theorem 1.4.

1.6.1 Measure-kernel-function

Theorem 1.20. Let K be a transition kernel from (E, \mathcal{E}) into (F, \mathcal{F}) . Then

$$Kf(x) = \int_F K(x, dy)f(y)$$

defines a function $Kf \in \mathcal{E}_+$ for every $f \in \mathcal{F}_+$;

$$\mu K(B) = \int_E \mu(dx)K(x, B)$$

defines a measure μK on (F, \mathcal{F}) for each measure μ on (E, \mathcal{E}) ; and

$$(\mu K)f = \mu(Kf) = \int_E \mu(dx) \int_F K(x, dy)f(y)$$

for every measure μ on (E, \mathcal{E}) and function f in \mathcal{F}_+ .

Proof. Kf is well-defined and measurable follows from theorem 1.4.

Define $L : \mathcal{F}_+ \rightarrow \overline{\mathbb{R}}_+ = f \mapsto \mu(Kf)$. Check

- $f(0) \Rightarrow L(f) = 0$

- If $f, g \in \mathcal{F}_+$ and $a, b \in \overline{\mathbb{R}}_+$, then:

$$\begin{aligned}
 L(af + bg) &= \mu(K(af + bg)) \\
 &\stackrel{\text{Linearity}}{=} \mu(aKf + bKg) \\
 &\stackrel{\text{linearity}}{=} a\mu(Kf) + b\mu(Kg) \\
 &= aL(f) + bL(g)
 \end{aligned}$$

- Suppose $(f_n) \subset \mathcal{F}_+$ and $f_n \nearrow f$, then

$$L(f_n) = \mu(Kf_n) \nearrow \mu(Kf) = L(f)$$

as MCT.

Hence, there exists a measure ν s.t. $L(f) = \mu(Kf) = \nu f$ as theorem 1.15. Then it suffices to show $\nu = \mu K$. Taking $f = \mathbf{1}_B$, we have $\nu(B) = \nu(\mathbf{1}_B) = \mu(K\mathbf{1}_B)$, it follows that

$$\mu(K\mathbf{1}_B) = \int_E \mu(dx) \int_F K(x, dy) \mathbf{1}_B(y) = \int_E \mu(dx) K(x, B) = \mu K(B)$$

□

Corollary 1.2. *A mapping $f \mapsto Kf : \mathcal{F}_+ \rightarrow \mathcal{E}_+$ specifies a transition kernel K iff*

- $K0 = 0$
- $K(af + bg) = aKf + bKg$ for $f, g \in \mathcal{F}_+$ and $a, b \in \overline{\mathbb{R}}_+$
- $Kf_n \nearrow Kf$ for every $(f_n) \nearrow f \subset \mathcal{F}_+$.

1.6.2 Products of kernels

Definition 1.17. Let K be a transition kernel from (E, \mathcal{E}) into (F, \mathcal{F}) and let L be a transition kernel from (F, \mathcal{F}) into (G, \mathcal{G}) . Then their **product** is the transition kernel KL from (E, \mathcal{E}) into (G, \mathcal{G}) defined by

$$(KL)f = K(Lf)$$

Remark. We can check KL is a transition kernel indeed by corollary 1.2. Obviously

$$KL(x, B) = \int_F K(x, dy) L(y, B)$$

1.6.3 Markov kernel

Definition 1.18. Let K be a transition kernel from (Ω, \mathcal{A}) into (Ω', \mathcal{A}') , it's called simply a transition kernel on (Ω, \mathcal{A}) if $\mathcal{A}' = \mathcal{A}$, moreover, it's called a **Markov kernel** if $K(x, \Omega) = 1, \forall x \in \Omega$ and a **sub-Markov kernel** if $K(x, \Omega) \leq 1, \forall x \in \Omega$.

If K is a transition kernel on (Ω, \mathcal{A}) , similarly with product kernel, we can define its **power** by $K^n = KK^{n-1}$ and $K^0 = I$ where I is the identity kernel on (Ω, \mathcal{A}) : $I(x, A) = \mathbf{1}_A(x)$. To see why it's called "identity", check

$$\begin{aligned} If(x) &= \int_{\Omega} I(x, dx) f(x) = \int_{\{x\}} f(x) = f(x) \\ \mu I(A) &= \int_{\Omega} \mu(dx) I(x, A) = \int_A \mu(dx) = \mu(A) \end{aligned}$$

and thus $IK = KI = K$. It follows that if K is Markov, so is K^n :

$$\begin{aligned} KK(x, \Omega) &= \int_{\Omega} K(x, dy) K(y, \Omega) \\ &= \int_{\Omega} K(x, dy) \\ &= K(x, \Omega) = 1 \end{aligned}$$

1.6.4 finite and bounded kernels

Definition 1.19. Let K be a transition kernel from (E, \mathcal{E}) into (F, \mathcal{F}) . In analogy with measures, it's called σ finite and finite if $B \mapsto K(x, B)$ is so for each $x \in E$

It's called bounded if $x \mapsto K(x, F)$ is bounded and σ bounded if there exists a partition $(F_n) \subset \mathcal{F}$ s.t. $x \mapsto K(x, F_n)$ is bounded for each n .

It's said to be s-finite if there exists countable finite (K_n) s.t. $K = \sum K_i$ and s-bounded if those (K_n) can be bounded.

If $K(x, \mathcal{F}) = 1$ for all x , the kernel is said to be a **transition probability kernel**.

Remark.

$$\begin{array}{ccccc} \text{bounded} & \implies & \sigma\text{-bounded} & \implies & s\text{-bounded} \\ \downarrow & & \downarrow & & \downarrow \\ \text{finite} & \implies & \sigma\text{-finite} & \implies & s\text{-finite} \end{array}$$

1.6.5 Functions on product spaces

Sections of a measurable function are measurable:

Proposition 1.2. *Let $f \in \mathcal{X} \times \mathcal{Y}$, then its selection, $x \mapsto f(x, y)$ and $y \mapsto f(x, y)$ are measurable for each x and y .*

Then we can generalize theorem 1.20 to functions on product spaces:

Lemma 1.11. *Let K be a s -finite kernel from (X, \mathcal{X}) into (X, \mathcal{Y}) , then, $\forall f \in (\mathcal{X} \times \mathcal{Y})_+$, define*

$$Tf(x) = \int_{\mathcal{Y}} f(x, y)K(x, dy) \in \mathcal{X}_+$$

moreover, $T : (\mathcal{X} \times \mathcal{Y}) \rightarrow \mathcal{X}_+$ is linear and continuous from below:

- $T(af + bg) = aTf + bTg$ for $f, g \in (\mathcal{X} \times \mathcal{Y})_+$ and $a, b \in \mathbb{R}_+$
- If $(f_n) \subset \mathcal{X} \times \mathcal{Y} \nearrow f$, then $Tf_n \nearrow Tf$.

Proof. By proposition 1.2, $f_x : y \mapsto f(x, y)$ is measurable in \mathcal{F}_+ and thus $Tf(x) = Kf_x(x)$, hence

- Linearity:

$$\begin{aligned} T(af + bg)(x) &= K(af_x + bg_x)(x) \\ &= aKf_x(x) + bKg_x(x) \\ &= aTf(x) + bTg(x) \\ &= (aTf + bTg)(x) \end{aligned}$$

- Continuity from below

$$f_n \nearrow f \implies Kf_{n_x}(x) \nearrow Kf_x(x) \implies Tf_n(x) \nearrow Tf(x)$$

Then it's remain to show $Tf \in \mathcal{X}_+$, assume K is bounded, suppose

$$\mathcal{M} = \{f \in (\mathcal{X} \times \mathcal{Y})_+ \cup (\mathcal{X} \times \mathcal{Y})_b : Tf \in \mathcal{X}\}$$

it's easy to check it's a monotone class and include all indicator of measurable rectangle $A \times B$. By theorem 1.5, we have $\mathcal{M} = (\mathcal{X} \times \mathcal{Y})_+ \cup (\mathcal{X} \times \mathcal{Y})_b$.

□

1.6.6 Measures on the product space

Theorem 1.21. *Let μ be a measure on (X, \mathcal{X}) and K be a s -finite kernel from (X, \mathcal{X}) to (Y, \mathcal{Y}) , then for any $f \in (\mathcal{X} \times \mathcal{Y})_+$*

$$\pi f = \int_X \int_Y f(x, y) K(x, dy) d\mu$$

define a measure π on the product space. Moreover, if μ is σ -finite and K is σ bounded, then π is σ finite and unique that satisfying:

$$\pi(A \times B) = \int_A K(x, B) d\mu$$

Proof. To see πf define a measure, check theorem 1.15, which follows from $\pi f = \mu(Tf)$ and similar properties enjoyed by T from lemma 1.11.

And the unique follows from theorem 1.8 by noting that all measurable rectangles is a π -system.

□

1.6.7 Product measures and Fubini

Definition 1.20. If $K(x, B) = \nu(B)$, i.e., independent to x , for some s -finite measure ν on (Y, \mathcal{Y}) , then such π is called **product** of μ and ν .

Theorem 1.22 (Fubini's theorem). *Let μ and ν be s -finite measures on (X, \mathcal{X}) and (Y, \mathcal{Y}) , respectively.*

- *There exists a unique s -finite measure π on $(X \times Y, \mathcal{X} \times \mathcal{Y})$ s.t. $\forall f \in (\mathcal{X} \times \mathcal{Y})_+$:*

$$\pi f = \int_X \int_Y f(x, y) d\nu d\mu = \int_Y \int_X f(x, y) d\mu d\nu$$

- *If $f \in \mathcal{X} \times \mathcal{Y}$ and $\pi f < \infty$, then $y \mapsto f(x, y)$ is ν integrable μ a.e. for every y , $x \mapsto f(x, y)$ is μ integrable ν a.e. for every x .*

Remark. For $A \in \mathcal{X}$ and $B \in \mathcal{Y}$, we have

$$\begin{aligned} \pi(A \times B) &= \pi \mathbf{1}_{A \times B} \\ &= \int_X \int_Y \mathbf{1}_{A \times B}(x, y) d\nu d\mu \\ &= \int_X \int_Y \mathbf{1}_A(x) \mathbf{1}_B(y) d\nu d\mu \\ &= \mu(A) \nu(B) \end{aligned}$$

and this is the reason we call π the product and write $\pi = \mu \times \nu$.

Remark. By theorem 1.21, only if both μ and ν are σ -finite the π is the unique product

1.6.8 Finite products

Now we can extend previous results to finitely many spaces' product. Similarly to product topology, $\prod_{i \in I} \mathcal{A}_i$ is generated by all measurable rectangles $\prod_{i \in I} A_i$ where I is finite.

Let (μ_n) be s-finite measures, their product measure is defined by analogy with theorem 1.22, $\forall f \in \prod_{i \in I} \mathcal{A}_i$,

$$\pi f = \int \dots \int f d\mu_n \dots d\mu_1$$

1.6.9 Infinite products

Similar again with product topology, $\prod_{i \in I} \mathcal{A}_i$ is generated by all measurable rectangles $\prod_{i \in I} A_i$ where $A_i = \Omega_i$ with finite exception. In analogy with topology product, we have:

Proposition 1.3. *Suppose there is $f_i : (X, \mathcal{F}) \rightarrow (\Omega_i, \mathcal{A}_i)$ for $i \in I$ and define $f(x) = (f_i(x))_{i \in I}$, then f is measurable iff each f_i is measurable.*

Chapter 2

Probability Spaces

2.1 Probability Spaces and Random Variables

Definition 2.1. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. The set Ω is called the **sample space** and whose elements are called **outcomes**. \mathcal{F} is called **history** and whose elements are called **events**.

Note here \mathbb{P} is finite measure, so it's continuous. We collect it's properties below :

Proposition 2.1. *For probability measure, which has following properties:*

1. $\forall A \in \mathcal{A}, \quad 0 \leq \mathbb{P}(A) \leq 1$
2. $\mathbb{P}(\Omega) = 1$
3. $\mathbb{P}(\sum_1^\infty A_n) = \sum_1^\infty \mathbb{P}(A_n)$
4. $\mathbb{P}(A) \leq \mathbb{P}(B) \iff A \subset B$
5. \mathbb{P} is continuous, as well as continuous from above and below.
6. **Boole's inequality**

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) \leq \sum_{i=1}^{\infty} \mathbb{P}(A_i)$$

2.1.1 Measure-theoretic and probabilistic languages

Measure	Probability
Integral	Expectation
Measurable set	Event
Measurable function	Random Variable
a.e.	a.s.

2.1.2 Distribution of a r.v.

Let X be a r.v. taking values in some measurable space (Y, \mathcal{Y}) , then let μ be the image of \mathbb{P} under X , i.e.:

$$\mu(A) = \mathbb{P}(X^{-1}A) = \mathbb{P}\{X \in A\}$$

then μ is a probability measure on (Y, \mathcal{Y}) , it's called the **distribution** of X . In view of theorem 1.8, it suffices to specify $\mu(A)$ for all A belongs to a π -system which generates \mathcal{Y} . In particular, if $(Y, \mathcal{Y}) = (\mathbb{R}, \mathcal{B})$, it's enough to specify

$$c(x) = \mu[-\infty, x] = \mathbb{P}\{X \leq x\}$$

and such $c : \mathbb{R} \rightarrow [0, 1]$ is called **distribution function(d.f.)**

Remark. Distribution function is nondecreasing and right continuous.

2.1.3 Joint distributions

Let X and Y taking values in (E, \mathcal{E}) and (F, \mathcal{F}) respectively then pair $Z = (X, Y)$ is measurable from \mathcal{F} to $\mathcal{E} \times \mathcal{F}$.

Recall the product spaces, to specifies distribution π of Z is suffices to:

$$\pi(A \times B) = \mathbb{P}\{X \in A, Y \in B\}$$

thus we have

$$\mu(A) = \mathbb{P}\{x \in A\} = \pi(A \times F)$$

μ and ν are called **marginal distributions**

2.1.4 Independence

Let X and Y taking values in (E, \mathcal{E}) and (F, \mathcal{F}) with marginal μ and ν , then they are said **independent** if their joint distribution is the product formed by their marginals:

$$\mathbb{P}\{X \in A, Y \in B\} = \mathbb{P}\{X \in A\}\mathbb{P}\{Y \in B\}$$

A finite collection $\{X_i\}_i^n$ is said to be **independency** if their product distribution has form $\prod_{i=1}^n \mu_i$. An arbitrary collection of r.v. is an independency if every finite subcollection is so.

2.1.5 Stochastic process and probability laws

Definition 2.2. Suppose $\{X_t : t \in T\}$ is a collection of r.v. taking values in (E, \mathcal{E}) . If T can be seen as time, then $(X_t)_{t \in T}$ is called a **stochastic process** with **state space** (E, \mathcal{E}) and **parameter set** T .

Now we can treat $X(\omega)$ as function $T \rightarrow E : t \mapsto X_t(\omega)$, thus $X : \mathcal{F} \rightarrow E^T$ is measurable as proposition 1.3 and it's a r.v. live in the same spaces as X_i and taking values in (E^T, \mathcal{E}^T) . It's distribution, $P \circ X^{-1}$ is called **probability law** of stochastic process $\{X_t : t \in T\}$.

Recall the product σ algebra construction, the probability law is determined by:

$$\mathbb{P}\left\{\bigcap_{i \in I} X_i \in A_i\right\}$$

where $I \subset T$ is finite and $A_i \subset E$

2.2 Expectation

Suppose X taking values in $\overline{\mathbb{R}}$, then we can talk about it's expectation:

$$\mathbb{E}X = \int_{\Omega} X d\mathbb{P} = \mathbb{P}X$$

the integral of X over an event $H \in \mathcal{F}$ is $\mathbb{E}X\mathbf{1}_H$

2.2.1 Properties of expectation

Suppose X, Y taking values in $\overline{\mathbb{R}}$ and $a, b > 0$. The following holds:

(Absolute integrability). $\mathbb{E}X$ is finite iff $\mathbb{E}|X|$ is finite.

(Positivity) If $X \geq 0$ a.s., then $\mathbb{E}X \geq 0$

(Monotonicity) If $X \geq Y$ or either $\mathbb{E}X$ and $\mathbb{E}Y$ is finite then both $\mathbb{E}X$ and $\mathbb{E}Y$ exist and $\mathbb{E}X \geq \mathbb{E}Y$.

(Linearity)

$$\mathbb{E}(aX + bY) = a\mathbb{E}X + b\mathbb{E}Y$$

(σ additivity over sets) If $A = \sum_{i=1}^{\infty} A_i$, then

$$\mathbb{E}_A X = \sum_{i=1}^{\infty} \mathbb{E}_{A_i} X$$

(Mean value theorem) If $a \leq X \leq b$ a.s., then

$$a\mathbb{P}(A) \leq \mathbb{E}_A X \leq b\mathbb{P}(A)$$

(Modulus inequality): $|\mathbb{E}X| \leq \mathbb{E}|X|$

(Fatou's) inequality If $X_n \geq 0$ a.s., then

$$\mathbb{E}\left(\liminf_n X_n\right) \leq \liminf_n \mathbb{E}X_n$$

(Dominated Convergence Theorem) If $X_n \rightarrow X$ a.s., $|X_n| \leq Y$ a.s. for all n and $\mathbb{E}Y < \infty$, then

$$\lim_n \mathbb{E}X_n = \mathbb{E}X = \mathbb{E}\lim_n X_n$$

(Monotone Convergence Theorem) If $0 \leq X_n \nearrow X$, then

$$\lim_n \mathbb{E}X_n = \mathbb{E}X = \mathbb{E}\lim_n X_n$$

(Integration term by term) If $\sum_{i=1}^{\infty} \mathbb{E}|X_n| < \infty$, then

$$\sum_{i=1}^{\infty} |X_n| < \infty, \text{ a.s.}$$

and

$$\mathbb{E}\left(\sum_{i=1}^{\infty} X_n\right) = \sum_{i=1}^{\infty} \mathbb{E}X_n$$

- Remark.*
1. If $\mathbb{P}(A) = 1$, then $\mathbb{E}X = \mathbb{E}_A X$.
 2. If $\mathbb{E}|X| < \infty$, then $|X| < \infty$ a.s., but not vice versa.
 3. If $X = Y$ a.s. and either $\mathbb{E}X$ or $\mathbb{E}Y$ exists, then so is the other and they are equal.
 4. $\forall H \in \mathcal{F}, \mathbb{E}X\mathbf{1}_H \geq \mathbb{E}Y\mathbf{1}_H \implies X \geq Y$ a.s. To see this, if there exist a subset $A \subset H$ s.t. $X < Y$ and $\mu(A) > 0$ then there is a contradiction with monotonicity in A .

2.2.2 Expectations and integrals

The following relates expectation and integrals w.r.t. distribution.

Theorem 2.1. *Let X be a r.v. taking value in (E, \mathcal{E}) then*

$$\int f \circ X d\mathbb{P} = \mathbb{E}f \circ X = \mu f = \int f d\mu$$

holds for all $f \in \mathcal{E}$ iff μ is the distribution of X .

Proof. Note $\mu = \mathbb{P} \circ X^{-1}$, then \Leftarrow follows from theorem 1.16. For \Rightarrow , taking $f = \mathbf{1}_A$:

$$\mu(A) = \mu \mathbf{1}_A = \mathbb{E} \mathbf{1}_A \circ X = \int \mathbf{1}_A \circ X d\mathbb{P}$$

that implies $\mu = \mathbb{P} \circ X^{-1}$ and claim follows. \square

Remark. By intuition, for a measure μ to be distribution of X it suffices to test all $f = \mathbf{1}_A$ for $A \in \mathcal{E}$ or even $A \in \mathcal{C}$ where \mathcal{C} is a π system and generating \mathcal{E} .

2.2.3 Means, variances, Laplace and Fourier transforms.

Definition 2.3. Let X be a r.v. taking values in $\overline{\mathbb{R}}$ with distribution μ , define

1. rth Moment: $\mathbb{E}X^r$
2. rth Absolute Moment: $\mathbb{E}|X|^r$
3. rth Central Moment: $\mathbb{E}(X - \mathbb{E}X)^r$
4. rth Absolute Central Moment: $\mathbb{E}|X - \mathbb{E}X|^r$
5. L^r space: $\{X : \mathbb{E}|X|^r < \infty\}$

Definition 2.4. Suppose $X \in \mathcal{F}_+$, for $r \in \mathbb{R}_+$, then $e^{-rX} \in [0, 1]$ and its expectation $\hat{\mu}_r = \mathbb{E}e^{-rX}$ also in $[0, 1]$. The resulting function $r \mapsto \hat{\mu}_r : \mathbb{R}_+ \rightarrow [0, 1]$ is called **Laplace transform** of the distribution μ , or Laplace transform of X for short.

Proposition 2.2. Let $X, Y \in \mathcal{F}_+$, TFAE:

1. X and Y have the same distribution
2. $\forall r \in \mathbb{R}_+, \mathbb{E}e^{-rX} = \mathbb{E}e^{-rY}$
3. $\mathbb{E}f \circ X = \mathbb{E}f \circ Y$ for every $f \in \mathbb{R}_c^{\mathbb{R}} \cap \mathbb{R}_b^{\mathbb{R}}$

Suppose that X is real-valued, for $r \in \mathbb{R}$, define:

$$\hat{\mu}_r = \mathbb{E}e^{irX} = \int (\cos rx + i \sin rx) d\mu$$

the resulting function $r \mapsto \hat{\mu}_r : \mathbb{R} \rightarrow \mathbb{C}$ is called the **Fourier transform** of μ or **characteristic function** of X

Proposition 2.3. Let $X, Y \in \mathcal{F}$, TFAE:

1. X and Y have the same distribution
2. $\forall r \in \mathbb{R}_+, \mathbb{E}e^{irX} = \mathbb{E}e^{irY}$
3. $\mathbb{E}f \circ X = \mathbb{E}f \circ Y$ for every $f \in \mathbb{R}_c^{\mathbb{R}} \cap \mathbb{R}_b^{\mathbb{R}}$

In particular, if $X \in \overline{\mathbb{N}}$, then for $z \in [0, 1]$, $\mathbb{E}z^X$ is called **generating function** and also determined distribution of X .

2.2.4 Moment inequalities

2.2.4.1 Young's inequality

Let f be a continuous and strictly increasing function with $f(0) = 0$, then we have

$$ab \leq \int_0^a f(x)dx + \int_0^b f^{-1}(x)dx$$

2.2.4.2 Holder's inequality

Suppose that $p, q > 1$ satisfy $\frac{1}{p} + \frac{1}{q} = 1$, then

$$E|XY| \leq [E|X|^p]^{1/p} [E|Y|^q]^{1/q}$$

2.2.4.3 Cauchy-Schwarz inequality

$$E|XY| \leq \sqrt{[E|X|^2][E|Y|^2]}$$

2.2.4.4 Lyapunov's inequality

1. $\forall p \geq 1, E|X| \leq E(|X|^p)^{\frac{1}{p}}$
2. $\forall 0 < r \leq s < \infty, [E|Z|^r]^{1/r} \leq [E|Z|^s]^{1/s}$

2.2.4.5 Minkowski's inequality

$\forall p \geq 1,$

$$(E|\sum X_i|^p)^{\frac{1}{p}} \leq \sum (E|X_i|^p)^{\frac{1}{p}}$$

2.2.4.6 Jensen's inequality

Let ψ be convex, that is, $\forall \lambda \in (0, 1), x, y \in \mathbb{R}$:

$$\lambda\psi(x) + (1 - \lambda)\psi(y) \geq \psi(\lambda x + (1 - \lambda)y)$$

Then

$$\psi(EX) \leq E[\psi(X)]$$

2.2.4.7 Chebyshev(Markov) inequality

If g is strictly increasing and positive on \mathbb{R}^+ , $g(x) = g(-x)$, and X is a r.v. s.t. $E(g(X)) < \infty$, then $\forall a > 0$

$$P(|X| \geq a) \leq \frac{Eg(X)}{g(a)}$$

Bonferroni's inequality

$$\mathbb{P} \left(\bigcap_{i=1}^n A_i \right) \geq \sum_{i=1}^n \mathbb{P}(A_i) - (n-1)$$