

Probability Notes 2024

Ruan Yuanlong

BUAA, BEIJING, CHINA

Email address: ruanyl@buaa.edu.cn

Contents

1. 单调类定理	6
2. 集函数与测度	12
2.1. 集函数	12
2.2. 半环上非负集函数	16
2.3. 环上非负集函数	22
3. Carathéodory's 延拓	27
3.1. 外测度	27
3.2. 域上测度的延拓	34
3.3. 半环上测度的延拓	39
3.4. Approximating $\mu^* _{\mathcal{F}_\mu^*}$ by $\mu^* _{\sigma(\mathcal{S})}$	41

3.5.	Approximating $\mu _{\sigma(\mathcal{A})}$ by $\mu _{\mathcal{A}}$	43
3.6.	Completion of a measure space	43
4.	收敛	48
4.1.	可测函数的收敛	48
4.2.	随机变量的分布函数	56
4.3.	随机变量的收敛	58
5.	积分	63
5.1.	非负可测函数积分	63
5.2.	可测函数积分	67
5.3.	Change of variables	72
6.	L_p 空间	72
6.1.	Inequalities	72
6.2.	Completeness	83
6.3.	L_p and weak convergence	86
6.4.	Uniform integrability	92
6.5.	Summary of various convergences	99
7.	概率空间的积分	100
7.1.	Expected value	100

7.2.	Properties of expectation	105
7.3.	Lebesgue-Stieltjes and Riemann-Stieltjes integrals	107
7.4.	L_p convergence and uniform integrability	111
8.	乘积测度空间	114
8.1.	Product σ -field	114
8.2.	Product measure space	117
8.3.	Fubini's Theorem	122
8.4.	Applications	124
8.5.	Finite-dimensional product space	129
9.	独立性 Independence	131
9.1.	Independence of events and random variables	131
9.2.	Independence and expectation	142
9.3.	Sum of independent random variables	147
10.	大数律 Law of large numbers	153
10.1.	L_2 weak law	153
10.2.	Weak law of large numbers	162

1. 单调类定理

Review:

- \mathcal{A} is a field, \mathcal{M} is a monotone class. Then

$$\mathcal{A} \subset \mathcal{M} \implies \sigma(\mathcal{A}) \subset \mathcal{M}.$$

- \mathcal{P} is a π -system, \mathcal{L} is a λ -system. Then

$$\mathcal{P} \subset \mathcal{L} \implies \sigma(\mathcal{P}) \subset \mathcal{L}.$$

- measurable spaces $(E, \mathcal{F}_E), (F, \mathcal{F}_F), f : (E, \mathcal{F}_E) \mapsto (F, \mathcal{F}_F)$.
 f is $\mathcal{F}_E/\mathcal{F}_F$ -measurable if

$$\sigma(f) \triangleq f^{-1}(\mathcal{F}_F) \subset \mathcal{F}_E.$$

Call it \mathcal{F}_E -measurable if

$$(F, \mathcal{F}_F) = (\mathbb{R}, \mathcal{B}(\mathbb{R})).$$

- $f : (E, \mathcal{F}_E) \mapsto (F, \sigma(\mathcal{E}))$, f is $\mathcal{F}_E/\sigma(\mathcal{E})$ -measurable if

$$f^{-1}(\mathcal{E}) \subset \mathcal{F}_E.$$

Thm 1 (π - λ theorem). \mathcal{P} is a π -system, \mathcal{L} is a λ -system. If $\mathcal{P} \subset \mathcal{L}$, then $\sigma(\mathcal{P}) \subset \mathcal{L}$.

Def 1 (Simple function). $i = 1, \dots, n$, $A_i \in \mathcal{F}$ (pairwise) disjoint, $c_i \in \mathbb{R}$. f is (measurable) simple if $f = \sum_{i=1}^n c_i 1_{A_i}$.

Alt. $i = 1, \dots, n$, $A_i \in \mathcal{F}$, $c_i \in \mathbb{R}$ non-zero distinct, f is simple if $f = \sum_{i=1}^n c_i 1_{A_i}$.

▷ 1. $a, b \in \mathbb{R}$, g simple, then $af + bg$ simple

Thm 2 (Simple approximation). (1) $f \geq 0$ measurable. There exist simple $\{f_n\}$, $0 \leq f_n \uparrow f$, uniform if f is bounded.

(2) f measurable. There exist simple $\{f_n\}$, $f_n \rightarrow f$, uniform if f is bounded.

PROOF. 1. Let

$$f_n = \frac{[2^n f]}{2^n} \wedge n = \sum_{i=0}^{n2^n-1} \frac{i}{2^n} 1_{\{i/2^n \leq f < (i+1)/2^n\}} + n 1_{\{f \geq n\}}.$$

Then

$$0 \leq f - f_n \leq \frac{1}{2^n} \text{ if } f < n; \quad f_n = n \leq f \text{ otherwise.}$$

2. $f = f^+ - f^-$. □

Thm 3 (Doob). $f : (E, \mathcal{F}_E) \mapsto (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, g measurable $(E, \mathcal{F}_E) \mapsto (F, \mathcal{F}_F)$. If f is $\sigma(g)$ -measurable, then $f = h \circ g$ for some measurable h .

PROOF. 1. $f = 1_A$, $A = g^{-1}(B) \in \sigma(g)$, $B \in \mathcal{F}_F$. Then $x \in A$ if and only if $g(x) \in B$, i.e.,

$$f = 1_A = 1_B \circ g.$$

2. f simple, $f = \sum_{i=1}^n c_i 1_{A_i}$, $c_i \in \mathbb{R}$, $A_i \in \sigma(g)$ disjoint. Let $A_i = g^{-1}(B_i)$, $B_i \in \mathcal{F}_F$, then

$$C_i = B_i \setminus \left(\bigcup_{j < i} B_j \right) \in \mathcal{F}_F \text{ disjoint}$$

and

$$f^{-1}(C_i) = A_i \setminus \left(\bigcup_{j < i} A_j \right) = A_i.$$

By step 1,

$$f = \sum_{i=1}^n c_i 1_{A_i} = \sum_{i=1}^n c_i 1_{C_i} \circ g = \left(\sum_{i=1}^n c_i 1_{C_i} \right) \circ g \triangleq h \circ g.$$

3. $f \geq 0$ is $\sigma(g)$ -measurable, there exist $\sigma(g)$ -measurable simple f_n with $0 \leq f_n \uparrow f$. It follows $f_n = h_n \circ g$ for some h_n ,

$$h \triangleq \sup_n h_n$$

is $\sigma(g)$ -measurable,

$$f = \lim_n f_n = \sup_n (h_n \circ g) = \left(\sup_n h_n \right) \circ g = h \circ g.$$

4. f is $\sigma(g)$ -measurable. f^+, f^- are $\sigma(g)$ -measurable. Use **3**. \square

Thm 4. \mathcal{A} is a π -system, $\Omega \in \mathcal{A}$, \mathcal{H} is a collection of real-valued functions. Suppose

(1) If $A \in \mathcal{A}$, then $1_A \in \mathcal{H}$

(2) If $f, g \in \mathcal{H}$, $c \in \mathbb{R}$, then $f + g, cg \in \mathcal{H}$

(3) If $f_n \in \mathcal{H}$, $0 \leq f_n \uparrow f$ with f bounded, then $f \in \mathcal{H}$

Then

$$\{f : f \text{ bounded } \sigma(\mathcal{A})\text{-measurable}\} \subset \mathcal{H}$$

PROOF. The system of sets

$$\mathcal{G} = \{A : 1_A \in \mathcal{H}\}$$

is a λ -system and $\mathcal{A} \subset \mathcal{G}$. Hence

$$\sigma(\mathcal{A}) \subset \mathcal{G}.$$

(2) implies that \mathcal{H} contains all $\sigma(\mathcal{A})$ -measurable simple functions, (3) implies that \mathcal{H} contains all bounded $\sigma(\mathcal{A})$ -measurable functions. \square

Thm 5. \mathcal{A} is a π -system, $\Omega \in \mathcal{A}$, \mathcal{H} is a collection of real-valued functions. Suppose

(1) If $A \in \mathcal{A}$, then $1_A \in \mathcal{H}$

(2) If $f, g \in \mathcal{H}$, $a, b \geq 0$, then $af + bg \in \mathcal{H}$

(3) If $f, g \in \mathcal{H}$ are bounded, $f \geq g$, then $f - g \in \mathcal{H}$

(4) If $f_n \in \mathcal{H}$, $0 \leq f_n \uparrow f$, then $f \in \mathcal{H}$

Then

$$\{f : f \text{ nonnegative } \sigma(\mathcal{A})\text{-measurable}\} \subset \mathcal{H}$$

2. 集函数与测度

2.1. 集函数. \mathcal{E} is a collection of subsets of E .

Def 2. *Set function, $\mu : \mathcal{E} \mapsto \mathbb{R} \cup \{\pm\infty\}$.*

Def 3. *Nonnegative set function, $\mu : \mathcal{E} \mapsto \mathbb{R} \cup \{\infty\}$.*

Def 4. *μ is finite if, $\forall A \in \mathcal{E}, |\mu(A)| < \infty$.*

Def 5. *μ is σ -finite on \mathcal{E} if, $\forall A \in \mathcal{E}$, there exist $\{A_n\} \subset \mathcal{E}$, $A = \bigcup_n A_n$ with $|\mu(A_n)| < \infty$.*

Def 6. *μ is additive if, $\forall A, B \in \mathcal{E}, AB = \emptyset$,*

$$\mu(A + B) = \mu(A) + \mu(B).$$

Def 7. *μ is countably additive if, $\forall A_i \in \mathcal{E}, i = 1, 2, \dots$, disjoint,*

$$\mu\left(\sum_i A_i\right) = \sum_i \mu(A_i).$$

Def 8. $\emptyset \in \mathcal{E}$. μ is a measure on \mathcal{E} if it is nonnegative, countably additive, $\mu(\emptyset) = 0$.

Example 1. (X, \mathcal{F}) measurable space, $x \in X$,

$$\delta_x(A) = 1_A(x), \quad \forall A \in \mathcal{F}.$$

$$x_1, \dots, x_n \in X,$$

$$\mu(A) = \sum_i \delta_{x_i}(A), \quad \forall A \in \mathcal{F}.$$

Example 2. F real-valued nonnegative, non-decreasing, right continuous. Semi-ring on \mathbb{R} ,

$$\mathcal{A} = \{(a, b] : a, b, \in \mathbb{R}\}.$$

Then

$$\mu((a, b]) = F(b) - F(a)$$

defines a measure \mathcal{A} . It is unique on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

PROOF. **1.** Additivity. $(a_i, b_i]$, $i = 1, \dots, n$, disjoint, $(a, b] = \bigcup_i^n (a_i, b_i]$, then

$$\mu((a, b]) = \sum_{i=1}^n \mu((a_i, b_i]).$$

2. $(a_i, b_i]$, $i = 1, \dots$, disjoint, $\bigcup_i (a_i, b_i] \subset (a, b]$, then

$$\sum_{i=1}^{\infty} \mu((a_i, b_i]) \leq \mu((a, b]).$$

3. $(a_i, b_i]$, $i = 1, \dots, n$, $(a, b] \subset \bigcup_i^n (a_i, b_i]$, then

$$\mu((a, b]) \leq \sum_{i=1}^n \mu((a_i, b_i]).$$

4. $(a_i, b_i]$, $i = 1, \dots$, disjoint, $\bigcup_i (a_i, b_i] = (a, b]$, then

$$\mu((a, b]) = \sum_{i=1}^{\infty} \mu((a_i, b_i]).$$

$\forall \varepsilon > 0$, there is $\delta_i > 0$,

$$F(b_i + \delta_i) - F(b_i) < \frac{\varepsilon}{2^i}.$$

$\forall \theta > 0$, $\{(a_i, b_i + \delta_i) : i\}$ is an open cover of $[a + \theta, b]$, there exists n_0

$$(a + \theta, b] \subset \bigcup_i^{n_0} (a_i, b_i + \delta_i].$$

By **3.**,

$$\begin{aligned}\mu((a + \theta, b]) &\leq \sum_{i=1}^{n_0} \mu((a_i, b_i + \delta_i]) \\ &= \sum_{i=1}^{n_0} (F(b_i + \delta_i) - F(b_i)) \\ &\leq \sum_{i=1}^{n_0} (F(b_i) - F(b_i)) + \sum_{i=1}^{n_0} \frac{\varepsilon}{2^i} \\ &\leq \sum_{i=1}^{\infty} (F(b_i) - F(b_i)) + \varepsilon.\end{aligned}$$

□

2.2. 半环上非负集函数. \mathcal{E} is a collection of subsets of E , μ is a nonnegative set function on \mathcal{E} .

Def 9. *Monotonicity:* $\forall A \subset B \in \mathcal{E}$,

$$\mu(A) \leq \mu(B).$$

Def 10. *Countably subadditive:* $\forall A_i \in \mathcal{E}, i = 1, 2, \dots, \bigcup_{i=1}^{\infty} A_i \in \mathcal{E}$,

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

Def 11. *Continuity from below:* $A_i \in \mathcal{E}, A_i \uparrow A \in \mathcal{E}$,

$$\lim_n \mu(A_i) = \mu(A).$$

Def 12. *Continuity from above:* $A_i \in \mathcal{E}, A_i \downarrow A \in \mathcal{E}, \mu(A_1) < \infty$,

$$\lim_n \mu(A_i) = \mu(A).$$

REMARK 1. **Note** finiteness is part of the definition of continuity from above.

\mathcal{S} is a semi-ring on E , μ is a nonnegative set function on \mathcal{S} .

Suppose μ is **additive**.

1. $\mu(\emptyset) = 0, +\infty$.

PROOF. $\emptyset \in \mathcal{S}$. By additivity

$$\mu(\emptyset) = \sum_{i=1}^n \mu(\emptyset).$$

$\mu(\emptyset)$ equals 0, or ∞ .

□

2. Monotonicity.

PROOF. $A, B \in \mathcal{S}$, $A \subset B$. There exist disjoint $C_1, \dots, C_k \in \mathcal{S}$,

$$B \setminus A = \bigcup_{i=1}^k C_i.$$

$$B = A \cup (B \setminus A) = A \cup \left(\bigcup_{i=1}^k C_i \right).$$

By additivity

$$\mu(B) = \mu(A) + \sum_{i=1}^k \mu(C_i) \geq \mu(A).$$

□

Suppose μ is **countably additive**.

3. Continuity from below.

PROOF. $A_i \in \mathcal{S}$, $A_i \uparrow A \in \mathcal{S}$. There exist disjoint $C_{n,1}, \dots, C_{n,k_n} \in \mathcal{S}$,

$$B_n \triangleq A_n \setminus A_{n-1} = \bigcup_{i=1}^{k_n} C_{n,i}.$$

$$(A_0 = \emptyset)$$

$$\begin{aligned}\mu(A) &= \mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \mu\left(\bigcup_{n=1}^{\infty} B_n\right) = \mu\left(\bigcup_{n=1}^{\infty} \bigcup_{i=1}^{k_n} C_{n,i}\right) \\ &= \sum_{n=1}^{\infty} \sum_{i=1}^{k_n} \mu(C_{n,i}) = \lim_N \sum_{n=1}^N \sum_{i=1}^{k_n} \mu(C_{n,i}) \\ &= \lim_N \mu\left(\bigcup_{n=1}^N \bigcup_{i=1}^{k_n} C_{n,i}\right) = \lim_n \mu(A_n).\end{aligned}$$

□

4. Continuity from above.

PROOF. (**WRONG PROOF**) $A_i \in \mathcal{S}$, $A_i \downarrow A \in \mathcal{S}$, $\mu(A_1) < \infty$.
Clearly

$$\mu\left(\bigcap_{n=1}^{\infty} A_n\right) \leq \mu(A_i) \leq \mu(A_1) < \infty.$$

$$\lim_n \mu(A_n) = \mu\left(\bigcap_{n=1}^{\infty} A_n\right)$$

$$\iff$$

$$\mu(A_1) - \lim_n \mu(A_n) = \mu(A_1) - \mu\left(\bigcap_{n=1}^{\infty} A_n\right)$$

$$\iff$$

$$\lim_n \mu(A_1 \setminus A_n) = \mu\left(A_1 \setminus \bigcap_{n=1}^{\infty} A_n\right) = \mu\left(\bigcup_{n=1}^{\infty} (A_1 \setminus A_n)\right).$$

□

5. Subadditivity.

PROOF. Analogous to continuity from below. □

2.3. 环上非负集函数.

Thm 6. \mathcal{R} is a ring. μ is nonnegative additive.

(1) μ countably additive



(2) μ countably subadditive



(3) μ continuity from below



(4) μ continuity from above



(5) μ continuity from above at \emptyset .

If μ is finite, (5) implies (1).

PROOF. **1.** Already have: $(1) \implies (2)$, $(1) \implies (3)$, $(1) \implies (4)$, $(4) \implies (5)$.

2. $(2) \implies (1)$. Suppose $A_i \in \mathcal{R}$, $i = 1, 2, \dots$, disjoint, $\bigcup_{i=1}^{\infty} A_i \in \mathcal{R}$.

By countable subadditivity,

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

By monotonicity and additivity,

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) \geq \mu\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n \mu(A_i), \quad \forall n.$$

Sending $n \rightarrow \infty$,

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) \geq \sum_{i=1}^{\infty} \mu(A_i).$$

3. (3) \implies (1). Suppose $A_i \in \mathcal{R}$, $i = 1, 2, \dots$, disjoint, $\bigcup_{i=1}^{\infty} A_i \in \mathcal{R}$.

Since

$$\bigcup_{i=1}^n A_i \uparrow \bigcup_{i=1}^{\infty} A_i,$$

by continuity from below,

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \lim_n \mu\left(\bigcup_{i=1}^n A_i\right) = \lim_n \sum_{i=1}^n \mu(A_i) = \sum_{i=1}^{\infty} \mu(A_i).$$

4. (5) \implies (1). Suppose $A_i \in \mathcal{R}$, $i = 1, 2, \dots$, disjoint, $\bigcup_{i=1}^{\infty} A_i \in \mathcal{R}$.

Then, $\forall n$,

$$\bigcup_{i=1}^n A_i \in \mathcal{R} \text{ and } \bigcup_{i=n+1}^{\infty} A_i = \bigcup_{i=1}^{\infty} A_i \setminus \bigcup_{i=1}^n A_i \in \mathcal{R}.$$

By additivity

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \mu\left(\bigcup_{i=1}^n A_i\right) + \mu\left(\bigcup_{i=n+1}^{\infty} A_i\right).$$

Since μ is finite

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) < \infty.$$

The continuity from above at \emptyset yields,

$$\lim_n \mu\left(\bigcup_{i=n+1}^{\infty} A_i\right) = 0.$$

Hence

$$\begin{aligned}\mu\left(\bigcup_{i=1}^{\infty} A_i\right) &= \lim_n \mu\left(\bigcup_{i=1}^n A_i\right) + \lim_n \mu\left(\bigcup_{i=n+1}^{\infty} A_i\right) \\ &= \lim_n \sum_{i=1}^n \mu(A_i) = \sum_{i=1}^{\infty} \mu(A_i).\end{aligned}$$

□

3. Carathéodory's 延拓

3.1. 外测度.

Def 13. μ^* is an outer measure on E if

(1) $\mu^*(\emptyset) = 0$

(2) $\forall A, B \in 2^E$, if $A \subset B$, then

$$\mu^*(A) \leq \mu^*(B)$$

(3) If $A_i \in 2^E, i = 1, 2, \dots$,

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

Thm 7. Let \mathcal{E} be a collection of sets on E , $\emptyset \in \mathcal{E}$. μ is a nonnegative set function on \mathcal{E} with $\mu(\emptyset) = 0$. Define, $\forall A \in 2^E$,

$$\mu^*(A) = \inf \left\{ \sum_{i=1}^{\infty} \mu(A_i) : A_i \in \mathcal{E}, A \subset \bigcup_{i=1}^{\infty} A_i \right\}.$$

Then $\mu^*(A)$ is an outer measure.

PROOF. **1.** $\mu^*(\emptyset) = 0$ since $\emptyset \in \mathcal{E}$, $\emptyset \subset \bigcup_{i=1}^{\infty} \emptyset$.

2. If $A \subset B$, $B \subset \bigcup_{i=1}^{\infty} B_i$, then $A \subset \bigcup_{i=1}^{\infty} B_i$, from the definition $\mu^*(A) \leq \mu^*(B)$.

3. Let $A_i \in 2^E, i = 1, 2, \dots, \varepsilon > 0$. There are $A_{i,k} \in \mathcal{E}$, $A_i \subset \bigcup_{k=1}^{\infty} A_{i,k}$,

$$\sum_{k=1}^{\infty} \mu(A_{i,k}) \leq \mu^*(A_i) + \frac{\varepsilon}{2^i}, \quad \forall i.$$

Since

$$\bigcup_{i=1}^{\infty} A_i \subset \bigcup_{i=1}^{\infty} \bigcup_{k=1}^{\infty} A_{i,k},$$

$$\begin{aligned}
\mu^*\left(\bigcup_{i=1}^{\infty} A_i\right) &\leq \sum_{i=1}^{\infty} \sum_{k=1}^{\infty} \mu(A_{i,k}) \\
&\leq \sum_{i=1}^{\infty} \left[\mu^*(A_i) + \frac{\varepsilon}{2^i} \right] \leq \sum_{i=1}^{\infty} \mu^*(A_i) + \varepsilon.
\end{aligned}$$

□

Def 14. μ^* is an outer measure on E . $A \in 2^E$ is μ^* -measurable if

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

The class of μ^* -measurable sets is denoted by \mathcal{F}_μ^* .

Def 15. Let μ be a measure on a σ -field \mathcal{F} of E , the measure space (E, \mathcal{F}, μ) is complete if

$$A \in \mathcal{F}, \quad \mu(A) = 0 \implies B \in \mathcal{F}, \quad \forall B \subset A.$$

Thm 8 (Carathéodory). *Let \mathcal{E} be a collection of sets on E , $\emptyset \in \mathcal{E}$. μ is a nonnegative set function on \mathcal{E} with $\mu(\emptyset) = 0$.*

(1) \mathcal{F}_μ^* is a σ -field.

(2) $(E, \mathcal{F}_\mu^*, \mu^*)$ is a complete measure space.

PROOF. 1. Obviously, $E \in \mathcal{F}_\mu^*$ and $A^c \in \mathcal{F}_\mu^*$ if $A \in \mathcal{F}_\mu^*$.

2. If $A_1, A_2 \in \mathcal{F}_\mu^*$, then $A_1 \cup A_2, A_1 \cap A_2 \in \mathcal{F}_\mu^*$.

$\forall D \in 2^E$, we note

$$D \cap (A_1 \cup A_2) = (D \cap A_1) \cup (D \cap A_1^c \cap A_2).$$

Then

$$\begin{aligned} & \mu^*(D \cap (A_1 \cup A_2)) + \mu^*(D \cap (A_1 \cup A_2)^c) \\ & \leq \mu^*(D \cap A_1) + \mu^*(D \cap A_1^c \cap A_2) + \mu^*(D \cap A_1^c \cap A_2^c) \quad (\text{subadditivity}) \\ & \leq \mu^*(D \cap A_1) + \mu^*(D \cap A_1^c) \quad (A_2 \in \mathcal{F}_\mu^*) \\ & = \mu^*(D) \quad (A_1 \in \mathcal{F}_\mu^*). \end{aligned}$$

Hence

$$A_1 \cup A_2 \in \mathcal{F}_\mu^*.$$

It follows that

$$(A_1 \cap A_2)^c = A_1^c \cup A_2^c \in \mathcal{F}_\mu^*.$$

3. Finite additivity. If $A_1, \dots, A_n \in \mathcal{F}_\mu^*$ disjoint, then $\forall D \in 2^E$,

$$\mu^* \left(D \cap \left(\bigcup_{i=1}^n A_i \right) \right) = \sum_{i=1}^n \mu^*(D \cap A_i).$$

Indeed, since $A_1 \in \mathcal{F}_\mu^*$,

$$\begin{aligned}
& \mu^* \left(D \cap \left(\bigcup_{i=1}^n A_i \right) \right) \\
&= \mu^* \left(D \cap \left(\bigcup_{i=1}^n A_i \right) \cap A_1 \right) + \mu^* \left(D \cap \left(\bigcup_{i=1}^n A_i \right) \cap A_1^c \right) \\
&= \mu^*(D \cap A_1) + \mu^* \left(D \cap \left(\bigcup_{i=2}^n A_i \right) \right) = \cdots = \sum_{i=1}^n \mu^*(D \cap A_i)
\end{aligned}$$

4. If $A_1, A_2, \dots \in \mathcal{F}_\mu^*$, then $A \triangleq \bigcup_{i=1}^{\infty} A_i \in \mathcal{F}_\mu^*$.

We can assume that $A_1, A_2, \dots \in \mathcal{F}_\mu^*$ are disjoint. Indeed, by **1** and **2**, $B_i = A_i \setminus \left(\bigcup_{j < i} A_j \right) \in \mathcal{F}_\mu^*$, are disjoint and $\bigcup_{i=1}^n B_i = \bigcup_{i=1}^n A_i$,

$\forall n$. Let

$$C_n = \bigcup_{i=1}^n A_i \in \mathcal{F}_\mu^*, \quad \forall n.$$

Since A_1, A_2, \dots are disjoint, we can use **3** (the finite additivity). $\forall D \in 2^E$,

$$\begin{aligned} \mu^*(D) &= \mu^*(D \cap C_n) + \mu^*(D \cap C_n^c) \\ &= \sum_{i=1}^n \mu^*(D \cap C_i) + \mu^*(D \cap C_n^c) \\ &\geq \sum_{i=1}^n \mu^*(D \cap C_i) + \mu^*(D \cap A^c), \quad \forall n. \end{aligned}$$

Let $n \rightarrow \infty$, note $A \subset \bigcup_{i=1}^{\infty} C_i$ and use subadditivity of outer measure

$$\mu^*(D) \geq \sum_{i=1}^{\infty} \mu^*(D \cap C_i) + \mu^*(D \cap A^c) \geq \mu^*(D \cap A) + \mu^*(D \cap A^c).$$

5. Countable additivity.

If $A_1, A_2, \dots, \in \mathcal{F}_\mu^*$ are disjoint, use **3** and send $n \rightarrow \infty$,

$$\mu^*\left(\bigcup_{i=1}^{\infty} A_i\right) \geq \mu^*\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n \mu^*(A_i), \quad \forall n.$$

The opposite inequality is subadditivity of outer measure.

6. Completeness. If $A \in \mathcal{F}_\mu^*$, $\mu^*(A) = 0$ and $B \subset A$, then $\mu^*(B) = 0$. $\forall D \in 2^E$,

$$\mu^*(D) \geq \mu^*(D \cap B^c) = \mu^*(D \cap B) + \mu^*(D \cap B^c).$$

So $B \in \mathcal{F}_\mu^*$. □

3.2. 域上测度的延拓.

Thm 9. *If μ is a measure on a field \mathcal{A} with the generated outer measure μ^* . Then*

(1) $\mathcal{A} \subset \mathcal{F}_\mu^*$ thus $\sigma(\mathcal{A}) \subset \mathcal{F}_\mu^*$.

(2) μ^* is an extension of μ to $\sigma(\mathcal{A})$ in the sense that

$$\mu(A) = \mu^*(A), \quad \forall A \in \mathcal{A}.$$

PROOF. 1. Let $A \subset \mathcal{A}$. If $A_i \in \mathcal{A}$, $A \subset \bigcup_{i=1}^{\infty} A_i$, then

$$(3.1) \quad \mu(A) \leq \sum_{i=1}^{\infty} \mu(A_i).$$

Indeed,

$$\mu\left(A \cap \bigcup_{i=1}^n A_i\right) \leq \mu\left(\bigcup_{i=1}^n A_i\right) \leq \sum_{i=1}^n \mu(A_i).$$

Let $n \rightarrow \infty$ and use that μ is a measure to get (3.1). So

$$\mu(A) \leq \mu^*(A).$$

Since $A \subset \mathcal{A}$, $A_1 = A$, $A_2 = A_3 \dots = \emptyset$ form a countable cover of A , so

$$\mu^*(A) \leq \mu(A).$$

2. Fix $A \subset \mathcal{A}$, will prove $A \in \mathcal{F}_\mu^*$. $\forall D \in 2^E$, it is enough to show that

$$\mu^*(D) \geq \mu^*(D \cap A) + \mu^*(D \cap A^c).$$

There is nothing to prove if $\mu^*(D) = \infty$, so we assume that $\mu^*(D) < \infty$. Then, $\forall \varepsilon > 0$, there exist $A_i \in \mathcal{A}$, $D \subset \bigcup_{i=1}^{\infty} A_i$ so that

$$\sum_{i=1}^{\infty} \mu(A_i) \leq \mu^*(D) + \varepsilon.$$

Since \mathcal{A} is a field,

$$A_i \cap A, A_i \cap A^c \in \mathcal{A}.$$

By **1** and the additivity of μ ,

$$\begin{aligned}\mu(A_i) &= \mu(A_i \cap A) + \mu(A_i \cap A^c) \\ &= \mu^*(A_i \cap A) + \mu^*(A_i \cap A^c).\end{aligned}$$

Summing over i gives

$$\begin{aligned}\sum_{i=1}^{\infty} \mu(A_i) &= \sum_{i=1}^{\infty} \mu^*(A_i \cap A) + \sum_{i=1}^{\infty} \mu^*(A_i \cap A^c) \\ &\geq \mu^*(D \cap A) + \mu^*(D \cap A^c).\end{aligned}$$

So

$$\mu^*(D) + \varepsilon \geq \sum_{i=1}^{\infty} \mu(A_i) \geq \mu^*(D \cap A) + \mu^*(D \cap A^c).$$

□

Thm 10 (Uniqueness). *Let \mathcal{P} be a π -system on E , μ and ν measures on $\sigma(\mathcal{P})$. Assume that*

(1) μ and ν agree on \mathcal{P} .

(2) There are $B_i \in \mathcal{P}$, $i = 1, 2, \dots$, disjoint so that $\bigcup_{i=1}^{\infty} B_i = E$ and

$$\mu(B_i) < \infty.$$

Then μ and ν are equal on $\sigma(\mathcal{P})$.

PROOF. 1. Let $B \in \mathcal{P}$ have $\mu(B) < \infty$. Define

$$\mathcal{L} = \{A \in \sigma(\mathcal{P}) : \mu(A \cap B) = \nu(A \cap B)\}.$$

\mathcal{L} is a λ -system (finiteness is needed to justify sets subtraction!), $\mathcal{P} \subset \mathcal{L}$. So

$$\sigma(\mathcal{P}) \subset \mathcal{L},$$

i.e.

$$\mu(A \cap B) = \nu(A \cap B), \quad \forall A \in \sigma(\mathcal{P}).$$

2. $\forall A \in \sigma(\mathcal{P})$, use (2) to write it as disjoint union,

$$A = \bigcup_{i=1}^{\infty} (A \cap B_i), \quad \mu(A \cap B_i) \leq \mu(B_i) < \infty.$$

Then, by 1,

$$\begin{aligned}\mu(A) &= \mu\left(\bigcup_{i=1}^{\infty}(A \cap B_i)\right) = \sum_{i=1}^{\infty} \mu(A \cap B_i) \\ &= \sum_{i=1}^{\infty} \nu(A \cap B_i) = \nu\left(\bigcup_{i=1}^{\infty}(A \cap B_i)\right) = \nu(A).\end{aligned}$$

□

▷ 2. The condition Theorem 10 (2) can be replaced with either one of the following:

(2') \mathcal{P} is a semi-ring, $E \in \mathcal{P}$ and μ is σ -finite on \mathcal{P} .

(2'') there are $B_1, B_2, \dots \in \mathcal{P}$, so that $B_i \uparrow E$ and $\mu(B_i) < \infty$.

3.3. 半环上测度的延拓.

Thm 11. Let μ be a measure on the semi-ring \mathcal{S} with the generated outer measure μ^* . Then

(1) $\mathcal{S} \subset \mathcal{F}_{\mu}^*$ thus $\sigma(\mathcal{S}) \subset \mathcal{F}_{\mu}^*$.

(2) μ^* is an extension of μ to $\sigma(\mathcal{S})$ in the sense that

$$(3.2) \quad \mu(A) = \mu^*(A), \quad \forall A \in \mathcal{S}.$$

(3) Assume that there are $B_i \in \mathcal{S}$, $i = 1, 2, \dots$, disjoint so that $\bigcup_{i=1}^n B_i = E$ and $\mu(B_i) < \infty$, then the extension of μ to $\sigma(\mathcal{S})$ is unique.

PROOF. Let $\bar{\mu}$ be the outer measure generated by μ .

1. $\bar{\mu}$ agrees with μ on \mathcal{S} .

The proof is identical to Theorem 9 (1).

2. Fix $A \subset \mathcal{S}$, will prove $A \in \mathcal{F}_\mu^*$.

The proof is identical to Theorem 9 (2). The difference is $A_i \cap A^c$ is replaced with disjoint union of sets in \mathcal{S} .

3. Uniqueness. Apply Theorem 10 to conclude. □

3.4. Approximating $\mu^*|_{\mathcal{F}_\mu^*}$ by $\mu^*|_{\sigma(\mathcal{S})}$.

Thm 12. *Let μ be a measure on the semi-ring \mathcal{S} with the generated outer measure μ^* . Suppose $E \in \mathcal{S}$.*

(1) *$\forall A \in \mathcal{F}_\mu^*$, there is $B \in \sigma(\mathcal{S})$ such that $A \subset B$ and*

$$\mu^*(A) = \mu^*(B).$$

(2) *If μ is σ -finite on \mathcal{S} , then $\forall A \in \mathcal{F}_\mu^*$, there is $B \in \sigma(\mathcal{S})$ such that $A \subset B$ and*

$$\mu^*(B \setminus A) = 0.$$

PROOF.

1. There is nothing to prove if $\mu^*(A) = \infty$, we assume that $\mu^*(A) < \infty$. There are $B_{n,i} \in \mathcal{S}$, $A \subset \bigcup_{i=1}^{\infty} B_{n,i}$,

$$\sum_{i=1}^{\infty} \mu(B_{n,i}) < \mu^*(A) + \frac{1}{n}.$$

Set

$$B = \bigcap_{n=1}^{\infty} \bigcup_{i=1}^{\infty} B_{n,i}.$$

Then $A \subset B \in \sigma(\mathcal{S})$,

$$\mu^*(A) \leq \mu^*(B).$$

Moreover

$$\mu^*(B) \leq \mu^*\left(\bigcup_{i=1}^{\infty} B_{n,i}\right) \leq \sum_{i=1}^{\infty} \mu(B_{n,i}) \leq \mu^*(A) + \frac{1}{n}.$$

It follows that

$$\mu^*(B) \leq \mu^*(A).$$

2. If μ is *finite* on \mathcal{S} , then by **1**, $\forall A \in \mathcal{F}_{\mu}^*$, there is $B \in \sigma(\mathcal{S})$ such that $A \subset B$ and

$$\mu^*(A) = \mu^*(B).$$

Since μ^* is a measure on \mathcal{F}_{μ}^* , this gives

$$\mu^*(B \setminus A) = 0.$$

The σ -finite case follows from similar argument as in step **3** of Theorem 11. \square

3.5. Approximating $\mu|_{\sigma(\mathcal{A})}$ by $\mu|_{\mathcal{A}}$.

Thm 13. *Let μ be a measure on the field \mathcal{A} with the generated outer measure μ^* . For any $A \in \sigma(\mathcal{A})$ with $\mu^*(A) < \infty$, $\forall \varepsilon > 0$, there is $B \in \mathcal{A}$ such that $\mu^*(A \Delta B) < \varepsilon$.*

If, in the last Theorem, the measure μ is defined on $\sigma(\mathcal{A})$ and σ -finite on \mathcal{A} , then μ must equal μ^* on $\sigma(\mathcal{A})$ by uniqueness, we can use μ in place of μ^* in the conclusion.

Thm 14. *Let \mathcal{A} be a field, μ a measure on $\sigma(\mathcal{A})$ and σ -finite on \mathcal{A} . For any $A \in \sigma(\mathcal{A})$ with $\mu(A) < \infty$, $\forall \varepsilon > 0$, there is $B \in \mathcal{A}$ such that $\mu(A \Delta B) < \varepsilon$.*

3.6. Completion of a measure space.

Thm 15. *Let (X, \mathcal{F}, μ) be a measure space,*

$$\bar{\mathcal{F}} \triangleq \{A \cup N : A \in \mathcal{F}, N \subset B \text{ for some } B \in \mathcal{F} \text{ with } \mu(B) = 0\}.$$

Define

$$\bar{\mu}(A \cup N) = \mu(A), \quad \forall A \in \bar{\mathcal{F}}.$$

Then $(X, \bar{\mathcal{F}}, \bar{\mu})$ is a complete measure space.

Clearly the Theorem says

$$\bar{\mu}(A) = \mu(A), \quad \forall A \in \bar{\mathcal{F}}.$$

PROOF. 1. $\bar{\mathcal{F}}$ is a σ -field.

Suppose $A \cup N \in \bar{\mathcal{F}}$ where $A \in \mathcal{F}$, $N \subset B$, $B \in \mathcal{F}$ with $\mu(B) = 0$.
Then

$$(A \cup N)^c = (A^c \cap B^c) \cup (B \cap A^c \cap N^c) \in \bar{\mathcal{F}}.$$

Suppose $A_i \cup N_i \in \bar{\mathcal{F}}$ where $A_i \in \mathcal{F}$, $N_i \subset B_i$, $B_i \in \mathcal{F}$ with $\mu(B_i) = 0$. Then

$$\bigcup_{i=1}^{\infty} (A_i \cup N_i) = \left(\bigcup_{i=1}^{\infty} A_i \right) \cup \left(\bigcup_{i=1}^{\infty} N_i \right) \in \bar{\mathcal{F}},$$

since

$$\bigcup_{i=1}^{\infty} N_i \subset \bigcup_{i=1}^{\infty} B_i \in \mathcal{F}$$

and

$$\mu\left(\bigcup_{i=1}^{\infty} B_i\right) = \sum_{i=1}^{\infty} \mu(B_i) = 0.$$

2. The definition of $\bar{\mu}$ nonambiguous, i.e.

$$A_1 \cup N_1 = A_2 \cup N_2 \in \tilde{\mathcal{F}} \implies \bar{\mu}(A_1 \cup N_1) = \bar{\mu}(A_2 \cup N_2).$$

Here $N_i \subset B_i$ for some $B_i \in \mathcal{F}$ with $\mu(B_i) = 0$, $i = 1, 2$.

$$\bar{\mu}(A_1 \cup N_1) = \mu(A_1) = \mu(A_1 \cup B_1 \cup B_2) \geq \mu(A_2) = \bar{\mu}(A_2 \cup N_2).$$

By symmetry,

$$\bar{\mu}(A_1 \cup N_1) \leq \bar{\mu}(A_2 \cup N_2).$$

(In fact

$$A_1 \cup B_1 \cup B_2 = A_1 \cup N_1 \cup B_1 \cup B_2 = A_2 \cup N_2 \cup B_1 \cup B_2 = A_2 \cup B_1 \cup B_2$$

so

$$\mu(A_1 \cup B_1 \cup B_2) = \mu(A_2).$$

)

3. Countable additivity. Suppose $A_i \cup N_i \in \bar{\mathcal{F}}$ disjoint, where $A_i \in \mathcal{F}$, $N_i \subset B_i$, $B_i \in \mathcal{F}$ with $\mu(B_i) = 0$. Then

$$\bar{\mu}\left(\bigcup_{i=1}^{\infty} (A_i \cup N_i)\right) = \mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i) = \sum_{i=1}^{\infty} \bar{\mu}(A_i \cup N_i).$$

4. Completeness. Let $A \cup N \in \bar{\mathcal{F}}$, $N \subset B$, $B \in \mathcal{F}$ with $\mu(B) = 0$ and $\bar{\mu}(A \cup N)$, then

$$\mu(A \cup B) = \mu(A) = \bar{\mu}(A \cup N) = 0.$$

So for any $C \subset A \cup N$, $C \subset A \cup B$,

$$C = \emptyset \cup C \in \bar{\mathcal{F}}.$$

□

Thm 16. *Suppose that μ is σ -finite on the semi-ring \mathcal{S} with the generated outer measure μ^* . Then $(X, \mathcal{F}_\mu^*, \mu^*)$ is the completion of $(X, \sigma(\mathcal{S}), \mu^*)$.*

PROOF. Let

$$\bar{\mathcal{F}} \triangleq \{A \cup N : A \in \sigma(\mathcal{S}), N \subset B \text{ for some } B \in \sigma(\mathcal{S}) \text{ with } \mu(B) = 0\}.$$

It is enough to show that

$$\mathcal{F}_\mu^* = \bar{\mathcal{F}}.$$

Since $(X, \mathcal{F}_\mu^*, \mu^*)$ is a complete measure space,

$$\bar{\mathcal{F}} \subset \mathcal{F}_\mu^*.$$

Let $A \in \mathcal{F}_\mu^*$, by Theorem 12 there exist $B, C \in \sigma(\mathcal{S})$ so that

$$A \subset B, \mu^*(B \setminus A) = 0; B \setminus A \subset C, \mu^*(C) = \mu^*(B \setminus A) = 0.$$

Writing

$$A = (B \cap C^c) \cup (A \cap C),$$

we get that $B \cap C^c \in \sigma(\mathcal{S})$, $(A \cap C) \subset C$, $\mu^*(C) = 0$, so $A \in \bar{\mathcal{F}}$. \square

4. 收敛

4.1. 可测函数的收敛. (E, \mathcal{F}, μ) a measure space, $f_n \in \mathcal{F}$, $i = 1, 2, \dots$, $f \in \mathcal{F}$

Def 16. *Almost everywhere convergence*, $f_n \xrightarrow{a.e.} f$:

$$\mu\left(\lim_n f_n \neq f\right) = 0.$$

Def 17. *Convergence in measure*, $f_n \xrightarrow{\mu} f$: $\forall \varepsilon > 0$,

$$\lim_n \mu(|f_n - f| > \varepsilon) = 0.$$

Evidently

$$\begin{aligned} f_n \xrightarrow{a.e.} f &\iff \forall \varepsilon > 0, \mu\left(\bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} \{|f_m - f| > \varepsilon\}\right) = 0 \\ &\iff \forall \varepsilon > 0, \mu(\{|f_n - f| > \varepsilon\} \text{ i.o.}) = 0. \end{aligned}$$

Recall

$$x \in \limsup A_n \iff x \in A_n \text{ i.o.}$$

Thm 17. *If μ is finite, then*

$$f_n \xrightarrow{a.e.} f \implies f_n \xrightarrow{\mu} f.$$

PROOF. Indeed,

$$\mu(|f_n - f| > \varepsilon) \leq \mu\left(\bigcup_{m=n}^{\infty} \{|f_m - f| > \varepsilon\}\right), \quad \forall n.$$

Let $n \rightarrow \infty$ and use continuity from above (requires finiteness of μ)

$$\begin{aligned} \limsup_n \mu(|f_n - f| > \varepsilon) &\leq \lim_n \mu\left(\bigcup_{m=n}^{\infty} \{|f_m - f| > \varepsilon\}\right) \\ &= \mu\left(\bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} \{|f_m - f| > \varepsilon\}\right) = 0. \end{aligned}$$

(or use

$$\limsup_n \mu(A_n) \leq \mu\left(\limsup_n A_n\right).$$

)

□

Def 18. *Almost uniform convergence, $f_n \xrightarrow{a.u.} f$: $\forall \varepsilon > 0$, there is $A_\varepsilon \in \mathcal{F}$ so that $\mu(A_\varepsilon) < \varepsilon$,*

$$\lim_n \sup_{x \notin A_\varepsilon} |f_n - f| = 0.$$

Compare with Egoroff's Theorem on *finite* measure!

Thm 18. $f_n \xrightarrow{a.u.} f$ if and only if $\forall \varepsilon > 0$,

$$\lim_n \mu \left(\bigcup_{m=n}^{\infty} \{|f_m - f| > \varepsilon\} \right) = 0.$$

PROOF. 1. " \implies ". $\forall \varepsilon > 0$, there is A_ε so that $\mu(A_\varepsilon) < \varepsilon$ and

$$\lim_m \sup_{x \notin A_\varepsilon} |f_m - f| = 0.$$

So, $\forall \varepsilon' > 0$, there is $n_0 \in \mathbb{N}$ such that

$$\sup_{x \notin A_\varepsilon} |f_m - f| \leq \varepsilon', \quad \forall m \geq n_0.$$

This translates to

$$\bigcup_{m=n_0}^{\infty} \{|f_m - f| > \varepsilon'\} \subset A_{\varepsilon}.$$

Therefore

$$\mu\left(\bigcup_{m=n_0}^{\infty} \{|f_m - f| > \varepsilon'\}\right) \leq \mu(A_{\varepsilon}) < \varepsilon.$$

2. " \Longleftarrow ". $\forall \varepsilon > 0$ and $k \in \mathbb{N}$, there is $n_{\varepsilon,k} \in \mathbb{N}$ such that

$$\mu\left(\bigcup_{m=n_{\varepsilon,k}}^{\infty} \left\{|f_m - f| > \frac{1}{k}\right\}\right) < \frac{\varepsilon}{2^k}, \quad \forall m \geq n_{\varepsilon,k}.$$

Denote (the set of all possible divergence points! measurable!)

$$A_{\varepsilon} = \bigcup_{k=1}^{\infty} \bigcup_{m=n_{\varepsilon,k}}^{\infty} \left\{|f_m - f| > \frac{1}{k}\right\}.$$

Then $\mu(A_\varepsilon) < \varepsilon$ and for any $x \notin A_\varepsilon$, we have $\forall k$,

$$|f_m - f| \leq \frac{1}{k}, \quad \forall m > n_{\varepsilon, k}.$$

□

We have proved:

Thm 19. (1)

$$f_n \xrightarrow{a.u.} f \implies f_n \xrightarrow{a.e.} f \text{ and } f_n \xrightarrow{\mu} f$$

(2) If μ is finite, then

$$f_n \xrightarrow{a.u.} f \iff f_n \xrightarrow{a.e.} f \implies f_n \xrightarrow{\mu} f$$

Example 3.

$$f_n(x) = \begin{cases} 1, & x \in (0, 1/n), \\ 0, & x \in [1/n, 1]. \end{cases}$$

Example 4.

$$f_n(x) = x^n, x \in [0, 1]$$

▷ 3. Let $f = 0$ and $f_n = 1_{A_n}$. Then $f_n \xrightarrow{\mu} f$ is equivalent to $\mu(A_n) \rightarrow 0$ and $\left(\lim_n f_n \neq f\right) = (A_n \text{ i.o.})$.

Any sequence $\{A_n\}$ so that $\mu(A_n) \rightarrow 0$ but $\mu(A_n \text{ i.o.}) > 0$ gives an example that $f_n \xrightarrow{\mu} f \not\Rightarrow f_n \xrightarrow{a.e.} f$. It is enough to have $\mu(A_n) \rightarrow 0$ and

$$\sum_{i=1}^{\infty} 1_{A_n}(x) = \infty, \quad \sum_{i=1}^{\infty} 1_{A_n^c}(x) = \infty.$$

Example 5. For each $n = 1, 2, \dots$ there is a unique decomposition $n = k(k-1)/2 + i$ with $k = 1, 2, \dots$, $i = 1, 2, \dots, k$.

$$f_n(x) = \begin{cases} 1, & x \in (((i-1)/k, i/k]), \\ 0, & \text{otherwise.} \end{cases}$$

Example 6. Consider

$$A_k^i = \left[\frac{i-1}{k}, \frac{i}{k} \right], \quad h_k^i(x) = 1_{A_k^i}(x), \quad i = 1, \dots, k.$$

Let f_n be the sequence

$$\{h_1^1; h_2^1, h_2^2; h_3^1, h_3^2; h_3^3; \dots\}$$

Thm 20. $f_n \xrightarrow{\mu} f \iff$ for any subsequence there is a further subsequence $f_{n_k} \xrightarrow{a.u.} f$.

PROOF. " \implies ". Since any subsequence of f_n converges in measure to f , it is enough to show there is a subsequence $f_{n_k} \xrightarrow{a.u.} f$. To see this, for any $k > 0$, by definition of convergence in measure, we can choose $n_k > n_{k-1}$ so that

$$\mu\left(|f_{n_k} - f| > \frac{1}{k}\right) \leq \frac{1}{2^k}.$$

Then

$$\mu\left(\bigcup_{k=m}^{\infty} |f_{n_k} - f| > \frac{1}{k}\right) \leq \sum_{k=m}^{\infty} \frac{1}{2^k} = \frac{1}{2^{m-1}}.$$

$\forall \varepsilon > 0$, for large m ,

$$\bigcup_{k=m}^{\infty} \{|f_{n_k} - f| > \varepsilon\} \subset \bigcup_{k=m}^{\infty} \left\{ |f_{n_k} - f| > \frac{1}{k} \right\}.$$

So

$$\lim_m \mu \left(\bigcup_{k=m}^{\infty} |f_{n_k} - f| > \varepsilon \right) \leq \lim_m \mu \left(\bigcup_{k=m}^{\infty} |f_{n_k} - f| > \frac{1}{k} \right) = 0.$$

” \Leftarrow ” Suppose $f_n \xrightarrow{\mu} f$ does not hold, i.e. there are $n_k \rightarrow \infty$, $\varepsilon_0 > 0$, $\delta_0 > 0$ so that

$$\mu(|f_{n_k} - f| > \varepsilon_0) > \delta_0.$$

Then

$$\liminf_m \mu \left(\bigcup_{k=m}^{\infty} |f_{n_k} - f| > \varepsilon_0 \right) \geq \delta_0,$$

Contradicting Theorem [18](#).

□

Theorem 19 and Theorem 20 indicate that if $f_n \xrightarrow{\mu} f$, then there is a subsequence $f_{n_k} \xrightarrow{a.e.} f$.

4.2. 随机变量的分布函数.

Def 19. (Ω, \mathcal{F}, P) is a probability space if P is a nonnegative measure on the σ -field \mathcal{F} with $P(\Omega) = 1$.

Def 20. A random variable (r.v.) X on (Ω, \mathcal{F}, P) is a real-valued mapping, $X : \omega \in \Omega \mapsto X(\omega) \in \mathbb{R}$.

Def 21. The distribution function of a r.v. X is

$$F(x) = P(X \leq x).$$

Denoted by $X \sim F$.

Thm 21. Any distribution function F has the following properties.

(1) non-decreasing, $F(-\infty) = 0$ and $F(\infty) = 1$

(2) right continuity: $\lim_{y \downarrow x} F(y) = F(x)$.

(3) left limit exists: $F(x-) = \lim_{y \uparrow x} F(y) = P(X < x)$.

$$(4) \quad P(X = x) = F(x) - F(x-).$$

The **inverse of the distribution function** F is defined as below.
 $\forall z \in (0, 1)$,

$$(4.1) \quad F^{-1}(z) = \inf\{x \in \mathbb{R} : F(x) \geq z\}.$$

▷ 4. Also equivalently defined as,

$$(4.2) \quad F^{-1}(z) = \sup\{x \in \mathbb{R} : F(x) < z\}.$$

LEMMA 22. F^{-1} has the properties,

- (1) F^{-1} is real-valued non-decreasing.
- (2) F^{-1} is left-continuous and has right limit.
- (3) $F^{-1}(F(x)) \leq x$, $F(F^{-1}(z)) \geq z$.
- (4) $F^{-1}(z) \leq x$ iff $F(x) \geq z$.

PROOF. Exercise. □

Thm 23. If F satisfies (1)(2)(3) of Theorem 21, there is a r.v. X with distribution F .

PROOF. Let $\Omega = (0, 1)$, $\mathcal{F} = \mathcal{B}_{(0,1)}$ (i.e. $(0, 1) \cap \mathcal{B}_{\mathbb{R}}$), $P =$ Lebesgue measure. Define

$$X(\omega) = F^{-1}(\omega).$$

Then X is \mathcal{F} -measurable (check this!) and

$$\begin{aligned} P(\omega : X(\omega) \leq x) &= P(\omega : F(x) \geq \omega) \\ &= \text{Lebesgue measure of } (0, F(x)) = F(x). \end{aligned}$$

So X is a r.v. with distribution function F . □

▷ 5. *Another construction of a r.v. X with distribution F is to take $\Omega = (\mathbb{R}, \mathcal{B})$, $P =$ the Lebesgue measure induced by F and consider the coordinate map $X(\omega) = \omega$.*

4.3. 随机变量的收敛. Probability space (Ω, \mathcal{F}, P) , r.v. X_n, X ,

$$X_n \xrightarrow{a.s.} X \iff P(X_n = X) = 1.$$

$$X_n \xrightarrow{P} X \iff \forall \varepsilon > 0, \lim_n P(|X_n - X| > \varepsilon) = 0.$$

Def 22. $X_n \sim F_n, X \sim F$. *Convergence in distribution (weak convergence):* $F_n(x) \rightarrow F(x)$ for all x where F is continuous, written $X_n \xrightarrow{d} X$.

Thm 24. $X_n \sim F_n, X \sim F$.

$$X_n \xrightarrow{a.s.} X \implies X_n \xrightarrow{P} X \implies X_n \xrightarrow{d} X.$$

PROOF. 1. The first implication is a special case of Theorem 17.

2. Check the second implication. $\forall \varepsilon, x \in \mathbb{R}, n \in \mathbb{N}$,

$$\begin{aligned} & P(X \leq x - \varepsilon) - P(|X_n - X| > \varepsilon) \\ & \leq P(X_n \leq x) \\ & \leq P(X_n \leq x, |X_n - X| \leq \varepsilon) + P(X_n \leq x, |X_n - X| > \varepsilon) \\ & \leq P(X \leq x + \varepsilon) + P(|X_n - X| > \varepsilon). \end{aligned}$$

So $n \rightarrow \infty, \varepsilon \rightarrow 0$ yield

$$F(x-) \leq \liminf_n P(X_n \leq x) \leq \limsup_n P(X_n \leq x) \leq F(x).$$

□

LEMMA 25. $F_n \xrightarrow{w} F \iff F_n^{-1} \xrightarrow{w} F^{-1}$.

PROOF OF " \implies ". Construct r.v.s' $X_n \sim F_n$, $X \sim F$ as Theorem 23. Fix any ω .

1. Choose any $\varepsilon > 0$ so that F is continuous at $X(\omega) - \varepsilon$ (the discontinuities of F are at most countable, ε can be arbitrarily small). By the definition (the infimum!) of $X(\omega)$,

$$F(X(\omega) - \varepsilon) < \omega.$$

Then, for large n ,

$$F_n(X(\omega) - \varepsilon) < \omega.$$

so (note the above inequality is strict)

$$X(\omega) - \varepsilon \leq X_n(\omega).$$

Hence

$$X(\omega) \leq \liminf_n X_n(\omega).$$

2. To see the opposite. Choose any $\varepsilon, \delta > 0$ so that X is continuous at ω and F is continuous at $X(\omega) + \varepsilon$, then by Lemma 22

$$F(X(\omega + \delta) + \varepsilon) \geq F(X(\omega + \delta)) \geq \omega + \delta > \omega.$$

For large n ($\delta > 0$),

$$F_n(X(\omega + \delta) + \varepsilon) \geq \omega.$$

By Lemma 22 again,

$$X(\omega + \delta) + \varepsilon \geq X_n(F_n(X(\omega + \delta) + \varepsilon)) \geq X_n(\omega).$$

Let $n \rightarrow \infty$, $\varepsilon \rightarrow 0$, $\delta \rightarrow 0$ (continuity at ω),

$$X(\omega) \geq \limsup_n X_n(\omega).$$

□

Thm 26 (Skorohod). $X_n \sim F_n$, $X \sim F$. Suppose $X_n \xrightarrow{d} X$. There exist r.v. \bar{X}_n, \bar{X} on a common probability space so that $\bar{X}_n \stackrel{d}{=} X_n$, $\bar{X} \stackrel{d}{=} X$, $\bar{X}_n \xrightarrow{a.s.} \bar{X}$.

PROOF. Let $\Omega = (0, 1)$, $\mathcal{F} = \mathcal{B}_{(0,1)}$, $P =$ Lebesgue measure. By Theorem 23 there exist r.v. on (Ω, \mathcal{F}, P) so that $\bar{X}_n \sim F_n$, $\bar{X} \sim F$. Lemma 25 then says $F_n^{-1} \xrightarrow{w} F^{-1}$. Since the discontinuity set of F^{-1} is countable, $F_n^{-1}(\omega) \rightarrow F^{-1}(\omega)$ for almost all $\omega \in \Omega$, i.e. $\bar{X}_n(\omega) \xrightarrow{a.s.} \bar{X}(\omega)$. \square

5. 积分

5.1. 非负可测函数积分. (E, \mathcal{F}, μ) a measure space, $f \in \mathcal{F}$ with values in $[0, \infty]$,. A *finite (measurable) partition* of E is a finite collection of \mathcal{F} -measurable sets $\{A_i : i = 1, \dots, m\}$ with $\bigcup_{i=1}^m A_i = E$.

$$(5.1) \quad \int f d\mu \triangleq \sup_{\text{finite partitions}} \sum_i \left[\inf_{x \in A_i} f(x) \right] \mu(A_i).$$

Convention: $0 \cdot \infty = 0$.

▷ 6. Consider

$$(5.2) \quad \int f d\mu \triangleq \inf_{\text{finite partitions}} \sum_i \left[\sup_{x \in A_i} f(x) \right] \mu(A_i).$$

Is (5.2) a good definition of integration?

Properties: $f, g \in \mathcal{F}$ nonnegative.

(1) If $f = 0$, μ -a.e., then $\int f d\mu = 0$.

(2) If $\mu(f > 0) > 0$, then $\int f d\mu > 0$.

(3) If $\int f d\mu < \infty$, then $f < \infty, \mu$ -a.e.

(4) If $f \leq g, \mu$ -a.e., then $\int f d\mu \leq \int g d\mu$.

(5) If $f = g, \mu$ -a.e., then $\int f d\mu = \int g d\mu$.

Thm 27 (Monotone convergence Theorem). *If $0 \leq f_n \uparrow f, \mu$ -a.e., then $0 \leq \int f_n d\mu \uparrow \int f d\mu$.*

PROOF. 1. First prove it under the assumption that

$$0 \leq f_n(x) \uparrow f(x), \forall x.$$

Integration is monotonic, so $\int f_n d\mu \leq \int f d\mu$. It remains to show

$$(5.3) \quad \lim_n \int f_n d\mu \geq \int f d\mu$$

or

$$\lim_n \int f_n d\mu \geq S = \sum_{i=1}^m c_i \mu(A_i)$$

for any finite measurable partition $\{A_i : i = 1, \dots, m\}$ and $c_i = \inf_{A_i} f$.

For such a partition, assume that the sum S , c_i and $\mu(A_i)$ are all finite. Fix $\alpha < 1$, define

$$A_{i,n} = \{x \in A_i : f_n(x) > \alpha c_i\}.$$

Since $f_n \uparrow f$, $A_{i,n} \uparrow A_i$. Consider the *measurable* partition

$$\{A_{i,n} : i = 1, \dots, m\} \cup \left\{ \left(\bigcup_{i=1}^m A_{i,n} \right)^c \right\}.$$

Then

$$\int f_n d\mu \geq \sum_{i=1}^m \alpha c_i \mu(A_{i,n}).$$

Let $n \rightarrow \infty$ and use continuity from below,

$$\lim_n \int f_n d\mu \geq \sum_{i=1}^m \alpha c_i \mu(A_i).$$

Finally let $\alpha \rightarrow 1$, (5.3) is proved.

Now suppose S is finite but not all of $c_i, \mu(A_i)$. Then $c_i \mu(A_i)$, $i = 1, \dots, m$ are finite. c_i or $\mu(A_i)$ may be infinity, but then $c_i \mu(A_i)$ must be zero. Use the adjusted partition $\{A_i : c_i \mu(A_i) > 0\} \cup \{\text{complement}\}$.

Lastly suppose S is infinite. Then there is some i_0 , $c_{i_0} \mu(A_{i_0}) = \infty$, i.e., $c_{i_0} > 0$, $\mu(A_{i_0}) > 0$ and at least one of them is ∞ . In this case

$$\int f d\mu = \infty.$$

To prove (5.3), let a, b satisfy

$$0 < a < c_{i_0} \leq \infty, \quad 0 < b < \mu(A_{i_0}) \leq \infty.$$

Define

$$A_{i_0, n} = \{x \in A_{i_0} : f_n(x) > a\}.$$

Since $f_n \uparrow f$, $A_{i_0,n} \uparrow A_{i_0}$ and $\mu(A_{i_0,n}) > b$ for n larger than some $n_{a,b}$. For the partition $\{A_{i_0,n}, A_{i_0,n}^c\}$, we have

$$\int f_n d\mu \geq a\mu(A_{i_0,n}) > ab, \forall n > n_{a,b}.$$

Let $a \rightarrow \infty$ if $c_{i_0} = \infty$, $b \rightarrow \infty$ if $\mu(A_{i_0,n}) = \infty$, we get

$$\lim_n \int f_n d\mu = \infty.$$

2. If $0 \leq f_n \uparrow f$ on A with $\mu(A^c) = 0$, then $0 \leq f_n 1_A \uparrow f 1_A$ holds everywhere. Then apply step **1**. \square

5.2. 可测函数积分. $f \in \mathcal{F}$ with values in $[-\infty, \infty]$,

$$\int f d\mu \triangleq \int f^+ d\mu - \int f^- d\mu.$$

f is said to be integrable if $\int f^+ d\mu, \int f^- d\mu$ are finite. So f integrable iff $|f|$ integrable.

Properties: $f, g \in \mathcal{F}$ integrable.

(1) If $f \leq g$, μ -a.e., then $\int f d\mu \leq \int g d\mu$.

(2) If $\alpha, \beta \in \mathbb{R}$, then $\alpha f + \beta g$ is integrable,

$$\int (\alpha f + \beta g) d\mu = \alpha \int f d\mu + \beta \int g d\mu.$$

Example 7. Let $E = \{1, 2, 3, \dots\}$, $\mathcal{F} = \{\text{all subsets of } E\}$, $\mu =$ counting measure. A function on E is a sequence x_1, x_2, \dots . Any function is \mathcal{F} -measurable. $\{x_k : k = 1, 2, \dots\}$ is μ -integrable if and only if $\sum_{k=1}^{\infty} |x_k|$ converges. When μ -integrable,

$$\sum_{k=1}^{\infty} |x_k| = \sum_{k=1}^{\infty} x_k^+ - \sum_{k=1}^{\infty} x_k^-.$$

The function $x_k = (-1)^{k+1}/k$, $k = 1, 2, \dots$ is not μ -integrable, although

$$\lim_m \sum_{k=1}^m (-1)^{k+1} \frac{1}{k} = \ln 2.$$

Thm 28 (Fatou's lemma). *Given f_n measurable.*

(1) *If g integrable, $f_n \geq g$, μ -a.e, then $\liminf_n f_n$ is integrable and*

$$\int \liminf_n f_n d\mu \leq \liminf_n \int f_n d\mu.$$

(1) *If g integrable, $f_n \leq g$, μ -a.e, then $\limsup_n f_n$ is integrable and*

$$\limsup_n \int f_n d\mu \leq \int \limsup_n f_n d\mu.$$

Thm 29 (Lebesgue's dominated convergence theorem). *Given g nonnegative integrable, $|f_n| \leq g$, μ -a.e.. If $f_n \xrightarrow{a.e.} f$ or $f_n \xrightarrow{\mu} f$, then*

$$\int f_n d\mu \longrightarrow \int f d\mu.$$

The following is a generalized dominated convergence theorem.

Thm 30. Given g_n nonnegative integrable, $|f_n| \leq g_n$, μ -a.e. with $g_n \xrightarrow{a.e.} g$ and $\int g_n d\mu \longrightarrow \int g d\mu$. If $f_n \xrightarrow{a.e.} f$ or $f_n \xrightarrow{\mu} f$, then

$$\int f_n d\mu \longrightarrow \int f d\mu.$$

Example 8 (Weierstrass M-test). If $|x_{n,m}| \leq M_m$, $\sum_{m=1}^{\infty} M_m < \infty$,

$\lim_n x_{n,m} = x_m$ for each m . Then

$$\lim_n \sum_{m=1}^{\infty} x_{n,m} = \sum_{m=1}^{\infty} x_m.$$

Example 9 (Bounded convergence theorem). Suppose μ is finite, $M > 0$. $|f_n| \leq M$, μ -a.e.. If $f_n \xrightarrow{a.e.} f$ or $f_n \xrightarrow{\mu} f$, then

$$\int f_n d\mu \longrightarrow \int f d\mu.$$

Example 10. If $f_n \geq 0$ or $\sum_{n=1}^{\infty} \int |f_n| d\mu < \infty$, then

$$\int \sum_{n=1}^{\infty} f_n d\mu = \sum_{n=1}^{\infty} \int f_n d\mu.$$

From this we get

Example 11. If $x_{n,m} \geq 0$ or $\sum_{m=1}^{\infty} \sum_{n=1}^{\infty} |x_{n,m}| < \infty$, then

$$\sum_{m=1}^{\infty} \sum_{n=1}^{\infty} x_{n,m} = \sum_{n=1}^{\infty} \sum_{m=1}^{\infty} x_{n,m}.$$

Example 12 (Abel's theorem). Suppose that the series $\sum_{k=1}^{\infty} |c_k|$ is convergent. Then

$$\lim_{x \rightarrow 1^-} \sum_{k=1}^{\infty} c_k x^k = \sum_{k=1}^{\infty} c_k.$$

5.3. Change of variables. (E_1, \mathcal{F}_1) , (E_2, \mathcal{F}_2) are measurable spaces, μ is a measure on \mathcal{F}_1 . T is measurable mapping from (E_1, \mathcal{F}_1) to (E_2, \mathcal{F}_2) . Define

$$(5.4) \quad \nu(B) = \mu(T^{-1}(B)), \quad \forall B \in \mathcal{F}_2.$$

Then $\nu(B)$ is a measure on \mathcal{F}_2 and for any $f \in \mathcal{F}_2$,

$$(5.5) \quad \int_{E_2} f d\nu = \int_{E_1} f \circ T d\mu.$$

Note if $f = 1_B$, then $f \circ T(x) = 1_B(T(x)) = 1_{T^{-1}(B)}(x)$, since $T(x) \in B$ iff $x \in T^{-1}(B)$. So in this case (5.5) reduces to (5.4).

6. L_p 空间

6.1. Inequalities.

LEMMA 31 (Jensen's inequality). *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space with $\mu(\Omega) = 1$, X a μ -integrable function on Ω , φ convex on \mathbb{R} . Then*

$$(6.1) \quad \varphi\left(\int_{\Omega} X d\mu\right) \leq \int_{\Omega} \varphi(X) d\mu.$$

Equality holds iff φ is linear on some convex set $A \subset \mathbb{R}$ with $\mu(X^{-1}A) = 1$.

PROOF. Denote by μ_X the induced measure of X on \mathbb{R} (ref section 5.3), then (6.1) is equivalent to

$$(6.2) \quad \varphi\left(\int_{\mathbb{R}} x d\mu_X\right) \leq \int_{\mathbb{R}} \varphi(x) d\mu_X$$

(Apply (5.5) with $f(x) = x$, $T = X$). It is enough to prove (6.2).

1. Denote $\bar{x} = \int_{\mathbb{R}} x d\mu_X$. Since φ is convex, there is a supporting line $L(x) = ax + b$ through \bar{x} , i.e. $L(\bar{x}) = \varphi(\bar{x})$ and

$$L(x) \leq \varphi(x), \quad \forall x.$$

Then

$$(6.3) \quad \int_{\mathbb{R}} L(x) d\mu_X \leq \int_{\mathbb{R}} \varphi(x) d\mu_X.$$

The LHS equals $\varphi\left(\int_{\mathbb{R}} x d\mu_X\right)$, hence (6.2) follows.

2. Suppose the equality in (6.2) holds, then by the above computation

$$\int_{\mathbb{R}} [\varphi(x) - L(x)] d\mu_X = 0.$$

The integrand is nonnegative, so the measurable set

$$A = \{x \in \mathbb{R} : \varphi(x) - L(x) = 0\}$$

has full measure, i.e. $\mu_X(A) = 1$. Moreover the set A is convex (verify directly!). On the other hand, if φ is linear on some convex $A \subset \mathbb{R}$ with $\mu(X^{-1}A) = 1$, then $\mu_X(A) = 1$,

$$\int_{\mathbb{R}} L(x) d\mu_X = \int_A L(x) d\mu_X, \quad \int_{\mathbb{R}} \varphi(x) d\mu_X = \int_A \varphi(x) d\mu_X.$$

Hence by (6.3),

$$\int_A [\varphi(X) - L(X)] d\mu \geq 0.$$

But the integrand $\varphi - L$ is nonnegative and linear on A . Since $A \subset \mathbb{R}$ is convex, it must be an interval. So the above integral is zero, hence the equality of (6.2) holds. \square

Notice that Lemma 31 does not require $\varphi(X)$ to be μ -integrable. From (6.3) it is clear that either $\int_{\Omega} \varphi(X) d\mu$ exists or equals infinity, in the latter case (6.1) trivially holds.

LEMMA 32. $a, b \in \mathbb{R}$, $1 \leq p < \infty$,

$$|a + b|^p \leq 2^{p-1}(|a|^p + |b|^p).$$

PROOF. Apply Jensen's inequality with $\varphi(x) = |x|^p$,

$$\left| \frac{a + b}{2} \right|^p \leq \frac{|a|^p + |b|^p}{2}.$$

□

LEMMA 33 (Young's inequality). $a, b \geq 0$, $1 < p, q < \infty$, $\frac{1}{p} + \frac{1}{q} = 1$,

$$a^{1/p} b^{1/q} \leq \frac{a}{p} + \frac{b}{q}.$$

Equal iff $a = b$.

PROOF. The inequality holds if $ab = 0$. In this case equality holds iff $a = b = 0$. Now suppose $ab > 0$. Apply Jensen's inequality with $\varphi(x) = -\ln x$,

$$-\ln\left(\frac{a}{p} + \frac{b}{q}\right) \leq -\frac{1}{p}\ln a - \frac{1}{q}\ln b.$$

Since φ is strictly convex (can touch a linear function at exactly one point), equality holds iff $a = b$. \square

(E, \mathcal{F}, μ) is a measure space in the following definitions.

Def 23. $p = 1$, let

$$L_1 \triangleq \{f \in \mathcal{F} : |f| \text{ is } \mu\text{-integrable}\}$$

and

$$\|f\|_1 = \|f\|_{L_1} = \int |f| d\mu.$$

Def 24. $1 < p < \infty$, let

$$L_p \triangleq \{f \in \mathcal{F} : |f|^p \in L_1\}$$

and

$$\|f\|_p = \|f\|_{L_p} = \left(\int |f|^p d\mu \right)^{1/p}.$$

Def 25. $p = \infty$, let

$$L_\infty \triangleq \{f \in \mathcal{F} : \text{there is } C > 0 \text{ such that } |f| \leq C, \text{ a.e.}\}$$

and

$$\|f\|_\infty = \|f\|_{L_\infty} = \inf\{C : |f| \leq C, \text{ a.e.}\}.$$

We could have written $L_p(\mu)$ to emphasize the dependence of the spaces L_p on the measure μ . But, when no ambiguity arises from the contexts, we will simply drop μ from the notation.

Thm 34 (Hölder inequality). $1 \leq p, q \leq \infty$, $\frac{1}{p} + \frac{1}{q} = 1$, $f \in L_p$, $g \in L_q$, then $fg \in L_1$ and

$$(6.4) \quad \|fg\|_1 \leq \|f\|_p \|g\|_q.$$

If $p = 1$, equality iff $|g| = \|g\|_\infty$, a.e. on the set where $f \neq 0$.

If $p = \infty$, equality iff $|f| = \|f\|_\infty$, a.e. on the set where $g \neq 0$.

If $1 < p < \infty$, equality iff there are nonnegative constants α, β such that $(\alpha, \beta) \neq (0, 0)$, $\alpha|f|^p = \beta|g|^q$, a.e.

PROOF. **1.** The inequality easily follows if $p = 1$ or $p = \infty$. To see the equality, suppose $p = 1$, then $q = \infty$. (6.4) is equivalent to

$$\int |f|(\|g\|_\infty - |g|) \geq 0.$$

It is equality iff $|g| = \|g\|_\infty$, a.e. on the set where $f \neq 0$.

2. Suppose $1 < p, q < \infty$. The conclusion is obvious if $\|f\|_p = 0$ or $\|g\|_q = 0$. Hence we assume that $0 < \|f\|_p, \|g\|_q < \infty$. Using Young's inequality with

$$a = \left(\frac{|f|}{\|f\|_p} \right)^p, \quad b = \left(\frac{|g|}{\|g\|_q} \right)^q,$$

we have

$$\frac{|fg|}{\|f\|_p\|g\|_q} \leq \frac{1}{p} \left(\frac{|f|}{\|f\|_p} \right)^p + \frac{1}{q} \left(\frac{|g|}{\|g\|_q} \right)^q, \text{ a.e.}$$

Integrating on both sides gives

$$\int \frac{|fg|}{\|f\|_p\|g\|_q} d\mu \leq \frac{1}{p} + \frac{1}{q} = 1,$$

which is the desired inequality. The equality holds iff $a = b$, *a.e.* i.e.,

$$\|g\|_q^q |f|^p = \|f\|_p^p |g|^q, \text{ a.e.}$$

□

A familiar case of Hölder inequality is the following.

Thm 35 (Cauchy–Schwarz inequality). *$f, g \in L_2$, then $fg \in L_1$ and*

$$\|fg\|_1 \leq \|f\|_2 \|g\|_2.$$

Thm 36 (Minkowski inequality). $1 \leq p \leq \infty$, $f, g \in L_p$, then $f + g \in L_p$ and

$$(6.5) \quad \|f + g\|_p \leq \|f\|_p + \|g\|_p.$$

If $p = 1$ or $p = \infty$, equality iff $fg \geq 0$, a.e..

If $1 < p < \infty$, equality iff there are nonnegative constants α, β such that $(\alpha, \beta) \neq (0, 0)$, $\alpha f = \beta g$, a.e.

PROOF. 1. The case $p = 1$ or $p = \infty$ is immediate.

2. Suppose $1 < p < \infty$. Let $q > 1$, $\frac{1}{p} + \frac{1}{q} = 1$. By Hölder inequality

$$\begin{aligned} \|f + g\|_p^p &= \int |f + g| |f + g|^{p-1} \leq_{(e1)} \int |f| |f + g|^{p-1} + \int |g| |f + g|^{p-1} \\ &\leq_{(e2)} \|f\|_p \| |f + g|^{p-1} \|_q + \|g\|_p \| |f + g|^{p-1} \|_q \\ &= \|f\|_p \|f + g\|_p^{p-1} + \|g\|_p \|f + g\|_p^{p-1} \end{aligned}$$

Here

$$\begin{aligned}\| |f + g|^{p-1} \|_q &= \left(\int (|f + g|^{p-1})^q \right)^{1/q} = \left(\int |f + g|^p \right)^{1/q} \\ &= \|f + g\|_p^{p/q} = \|f + g\|_p^{p-1}.\end{aligned}$$

(e1) is equality iff $fg \geq 0$, *a.e.*, (e2) is equality iff there are nonnegative constants a, b, c, d such that $(a, b) \neq (0, 0)$, $(c, d) \neq (0, 0)$,

$$a|f|^p = b(|f + g|^{p-1})^q, \quad c|g|^p = d(|f + g|^{p-1})^q, \quad \textit{a.e.}$$

Hence

$$a|f| = b|f + g|, \quad c|g| = d|f + g|, \quad \textit{a.e.}$$

The conclusion follows by combining the equality conditions of (e1)(e2). \square

Def 26. $0 < p < 1$, let

$$L_p \triangleq \left\{ f \in \mathcal{F} : \int |f|^p d\mu < \infty \right\}$$

and

$$\|f\|_p = \int |f|^p d\mu.$$

LEMMA 37. Let $a, b \in \mathbb{R}$, $0 < p < 1$. $|a + b|^p \leq |a|^p + |b|^p$.

PROOF. Since $||a| + |b||^p \leq |a|^p + |b|^p$ implies the desired inequality, we assume w.l.g. that a, b are of the same sign. Suppose $a \neq 0$, otherwise there is nothing to prove. Finally it suffices to show that

$$(1 + s)^p \leq 1 + s^p, \quad s \geq 0,$$

which is verified by elementary calculus. □

Lemma 32 and Lemma 37 can be merged into the compact form,

$$(6.6) \quad |a + b|^p \leq C_p(|a|^p + |b|^p), \quad 0 < p < \infty,$$

where $C_p = 2^{p-1} \vee 1$.

Thm 38. $0 < p < 1$, $\|f + g\|_p \leq \|f\|_p + \|g\|_p$.

6.2. Completeness.

Thm 39. *Let $0 < p \leq \infty$, L_p is complete.*

PROOF FOR $p = \infty$. Let $f_n \in L_\infty$. Suppose that f_n is Cauchy. Given $k \geq 1$, there is n_k such that

$$\|f_m - f_n\|_\infty \leq \frac{1}{k}, \quad \forall m, n > n_k.$$

Hence there is a null set¹ A_k such that

$$|f_m - f_n| \leq \frac{1}{k}, \quad \forall x \in A_k^c, \quad m, n > n_k.$$

Then $A = \bigcup_{k=1}^{\infty} A_k$ is a null set and $f_n(x)$ is Cauchy for each $x \in A^c$.

Hence there exist f , $f_n \rightarrow f$ for $x \in A^c$. Let $m \rightarrow \infty$ in the above inequality we get

$$|f_n - f| \leq \frac{1}{k}, \quad \forall x \in A^c, \quad n > n_k.$$

¹A null set is a measurable set with measure zero.

So $f \in L_\infty$ and

$$\|f_n - f\|_\infty \leq \frac{1}{k}, \quad \forall n > n_k.$$

Therefore f_n converges to f in L_∞ . □

PROOF FOR $0 < p < \infty$. Let $f_n \in L_p$. Suppose that f_n is Cauchy in L_p ,

$$(6.7) \quad \lim_{m, n \rightarrow \infty} \|f_m - f_n\|_p = 0.$$

We intend to show that $\lim_{n \rightarrow \infty} \|f_n - f\|_p = 0$ for some $f \in L_p$. Owing to (6.7), we have a subsequence $n_k \rightarrow \infty$ so that

$$(6.8) \quad \|f_{n_{k+1}} - f_{n_k}\|_p < \frac{1}{2^k}.$$

We claim that

- (a) there is $h \in L_p$ such that $|f_{n_k}| \leq h$, *a.e.*
- (b) $\lim_k f_{n_k} \rightarrow f$, *a.e.* for some $f \in L_p$.
- (c) $\lim_k \|f_{n_k} - f\|_p = 0$.

The conclusion of the Theorem clearly follows once (c) is proved, since a Cauchy sequence converges iff it has a convergent subsequence. Let

$$g_k = \sum_{i=1}^k |f_{n_{i+1}} - f_{n_i}|, \quad g = \sum_{i=1}^{\infty} |f_{n_{i+1}} - f_{n_i}|.$$

Then $0 \leq g_k \uparrow g$ and $\|g_k\|_p \leq 1$ by (6.8) (Theorem 36 or Theorem 38). Using monotone convergence theorem,

$$\int g^p d\mu = \lim_k \int (g_k)^p d\mu \leq 1.$$

This shows $g \in L_p$ and that $g < \infty$, *a.e.* Therefore

$$f_{n_k} = f_{n_1} + \sum_{i=1}^k (f_{n_{i+1}} - f_{n_i})$$

converges almost everywhere to some measurable function f and

$$|f_{n_k}| \leq |f_{n_1}| + g.$$

Let $k \rightarrow \infty$, we have

$$|f| \leq |f_{n_1}| + g, \text{ a.e.}$$

hence $f \in L_p$. (a)(b) follows with $h = |f_{n_1}| + g$. By inequality (6.6),

$$\begin{aligned} |f_{n_k} - f|^p &\leq C_p(|f_{n_k}|^p + |f|^p) \leq C_p(|f_{n_1}| + g)^p + |f|^p \\ &\leq C_p(C_p(|f_{n_1}|^p + g^p) + |f|^p). \end{aligned}$$

Therefore (c) is a result of the dominated convergence theorem. □

COROLLARY 1. (1) $0 < p < 1$, L_p is a complete metric space.
 (2) $1 \leq p \leq \infty$, L_p is a Banach space.

6.3. L_p and weak convergence.

Thm 40. Let $0 < p < \infty$, $f_n \in L_p$, $f \in L_p$.

- (1) $f_n \xrightarrow{L_p} f \implies f_n \xrightarrow{\mu} f$ and $\|f_n\|_p \rightarrow \|f\|_p$.
 (2) $f_n \xrightarrow{\text{a.e.}} f$ or $f_n \xrightarrow{\mu} f$, then

$$\|f_n\|_p \rightarrow \|f\|_p \iff f_n \xrightarrow{L_p} f.$$

PROOF. **1.** To prove (1), use Markov inequality

$$\mu(|f_n - f| > \varepsilon) \leq \frac{1}{\varepsilon^p} \|f_n - f\|_p^p.$$

and the triangle inequality

$$\left| \|f_n\|_p - \|f\|_p \right| \leq \|f_n - f\|_p.$$

2. " \Leftarrow " of (2) is included in step **1**.

3. " \Rightarrow " of (2). In view of Theorem [20](#), it is enough to prove the case where $f_n \xrightarrow{a.e.} f$. Define

$$g_n = C_p(|f_n|^p + |f|^p) - |f_n - f|^p,$$

where $C_p = 2^{p-1} \vee 1$. Then $g_n \geq 0$ by inequality (6.6) and $\lim_n g_n = 2C_p|f|^p$, a.e. Using Fatou's lemma

$$\begin{aligned} \int 2C_p|f|^p d\mu &= \int \lim_n g_n d\mu \leq \liminf_n \int g_n d\mu \\ &= \int 2C_p|f|^p d\mu - \limsup_n \int |f_n - f|^p. \end{aligned}$$

Canceling $\int 2C_p|f|^p d\mu$ from both side gives

$$\lim_n \int |f_n - f|^p = 0.$$

□

Def 27. (E, \mathcal{F}, μ) is a measure space. $1 \leq p < \infty$, $\frac{1}{p} + \frac{1}{q} = 1$, f_n converges weakly to f in L_p , denoted by $f_n \xrightarrow{w-L_p} f$, if

$$\lim_n \int f_n g d\mu = \int f g d\mu, \quad \forall g \in L_q.$$

μ is additionally assumed to be σ -finite if $p = 1$.

Thm 41. $1 \leq p < \infty$. $f_n \xrightarrow{L_p} f$ implies $f_n \xrightarrow{w-L_p} f$.

PROOF. By Hölder inequality (Theorem 34), $\forall g \in L_q$, q conjugate to p ,

$$\int |f_n - f| |g| d\mu \leq \|f_n - f\|_p \|g\|_q.$$

□

Thm 42. (E, \mathcal{F}, μ) is a measure space. Let $1 < p < \infty$, $\{f_n\}$ bounded in L_p . If $f_n \xrightarrow{a.e.} f$ or $f_n \xrightarrow{\mu} f$ for some measurable f , then $f \in L_p$ and $f_n \xrightarrow{w-L_p} f$.

PROOF. Let $g \in L_q$, q conjugate to p . As before, it is enough to prove it for $f_n \xrightarrow{a.e.} f$.

1. $f \in L_p$ is a consequence of Fatou's lemma,

$$\int |f|^p d\mu = \int \lim_n |f_n|^p d\mu \leq \liminf_n \int |f_n|^p d\mu \leq \sup_n \|f_n\|_{L_p}^p < \infty.$$

It follows that $\{f_n - f\}$ is bounded in L_p .

2. Fix $\varepsilon > 0$, let $\delta > 0$, define $A_\delta = \{x \in E : \delta \leq |g|^q \leq 1/\delta\}$ and write

$$\int |f_n - f| |g| d\mu = \int_{A_\delta \cap B} + \int_{A_\delta \cap B^c} + \int_{A_\delta^c}.$$

Choose δ small so that

$$\int_{A_\delta^c} \leq \|f_n - f\|_p \|g 1_{A_\delta^c}\|_q < \frac{\varepsilon}{3}.$$

With δ fixed, we have

$$\int_{A_\delta \cap B^c} \leq \|f_n - f\|_p \|g 1_{A_\delta \cap B^c}\|_q < \frac{\varepsilon}{3},$$

as soon as $B \subset A_\delta$ is such that $\mu(A_\delta \cap B^c)$ is smaller than some ε' .

Note $|g| \leq 1/\delta^{1/q}$ on A_δ . Since $\mu(A_\delta)$ is finite by Markov inequality, so a subset $B \subset A_\delta$ can be chosen so that $\mu(A_\delta \cap B^c) < \varepsilon'$ and $|f_n - f|$

converges uniformly to 0 on $A_\delta \cap B$ (Theorem 19). Hence for large n ,

$$\int_{A_\delta \cap B} \leq \frac{1}{\delta^{1/q}} \int_{A_\delta \cap B} |f_n - f| d\mu < \frac{\varepsilon}{3}.$$

□

Note the above proof does not get through if $p = 1$ (so that $q = \infty$). The example below demonstrates, in general, Theorem 42 does not for $p = 1$.

Example 13. $E = (0, 1)$ with the usual Lebesgue measure, $f_n = n1_{(0, 1/n)}$. Clearly $\|f_n\|_1 = 1$, $f_n \xrightarrow{\mu} f = 0$. But with $g = 1 \in L_\infty$, $\lim_n \int f_n g d\mu = 1 \neq 0 = \int f g d\mu$, hence $f_n \xrightarrow{w-L_1} f$ does not hold.

However we have

Thm 43. (E, \mathcal{F}, μ) is a measure space. Let $\{f_n\} \in L_1$. Suppose $f_n \xrightarrow{a.e.} f$ or $f_n \xrightarrow{\mu} f$. Then

$$f \in L_1, \quad \|f_n\|_1 \rightarrow \|f\|_1 \iff f_n \xrightarrow{L_1} f.$$

Either of them gives $\int_A f_n d\mu \rightarrow \int_A f d\mu, \forall A \in \mathcal{F}$.

PROOF. The first conclusion is contained in Theorem 40. So $f_n \xrightarrow{w-L^2} f$ by Theorem 41. To complete the proof, take $1_A \in L_\infty$ as test function. \square

6.4. Uniform integrability. Let (E, \mathcal{F}, μ) be a measure space.

Def 28. $\mathcal{H} = \{f_t : t \in T\}$ is uniformly integrable if

$$(6.9) \quad \lim_{a \rightarrow \infty} \sup_{f \in \mathcal{H}} \int_{\{|f| \geq a\}} |f| d\mu = 0.$$

Def 29. $\mathcal{H} = \{f_t : t \in T\}$ is absolutely continuous if, $\forall \varepsilon > 0$, there is $\delta > 0$ so that

$$\sup_{f \in \mathcal{H}} \int 1_A |f| d\mu < \varepsilon \text{ for any } A \text{ with } \mu(A) < \delta.$$

Thm 44. Suppose (E, \mathcal{F}, μ) is a measure space with μ finite. $\mathcal{H} = \{f_t : t \in T\}$ is uniformly integrable if and only if \mathcal{H} is absolutely continuous and bounded in L_1 .

PROOF. 1. If \mathcal{H} is uniformly integrable, $\forall \varepsilon > 0$, there is $a_0 > 0$ so that

$$\sup_{f \in \mathcal{H}} \int_{\{|f| \geq a\}} |f| d\mu \leq \frac{\varepsilon}{2}, \quad \forall a \geq a_0.$$

For any measurable A , $a \geq a_0$,

$$\begin{aligned} \sup_{f \in \mathcal{H}} \int 1_A |f| d\mu &\leq \sup_{f \in \mathcal{H}} \int_{\{|f| < a\}} 1_A |f| d\mu + \sup_{f \in \mathcal{H}} \int_{\{|f| \geq a\}} 1_A |f| d\mu \\ &\leq a\mu(A) + \sup_{f \in \mathcal{H}} \int_{\{|f| \geq a\}} |f| d\mu \leq a\mu(A) + \frac{\varepsilon}{2}. \end{aligned}$$

That \mathcal{H} is bounded in L_1 follows by setting $A = E$ and using the fact that μ is finite. Fix $a \geq a_0$. For any A with $\mu(A) \leq \varepsilon/(2a)$, we get that $\sup_{f \in \mathcal{H}} \int 1_A |f| d\mu$ is bounded from above by ε , hence the absolute continuity.

2. Suppose that \mathcal{H} is absolutely continuous and bounded in L_1 . Denote the uniform L_1 bound of \mathcal{H} by M . By Markov inequality, $\forall a > 0$,

$$\mu(|f| > a) \leq \frac{1}{a} \int |f| d\mu \leq \frac{1}{a} M, \quad \forall f \in \mathcal{H}.$$

$\forall \varepsilon > 0$, by absolute continuity, $\sup_{f \in \mathcal{H}} \int 1_A |f| d\mu < \varepsilon$ as soon as $\mu(A)$ is less than some $\delta > 0$. Fix a with $M/a < \delta$. Then setting $A = \mu(|f| > a)$ gives the uniform integrability. \square

Thm 45 (Vitali convergence theorem). *Suppose that μ is finite, $f_n \xrightarrow{a.e.} f$ or $f_n \xrightarrow{\mu} f$.*

(1) *If $\{f_n\}$ is uniformly integrable, then $f \in L_1$ and*

$$(6.10) \quad \int f_n d\mu \rightarrow \int f d\mu.$$

(2) *f_n, f are nonnegative integrable, then (6.10) implies that $\{f_n\}$ is uniformly integrable.*

PROOF. The proof is given for $f_n \xrightarrow{a.e.} f$.

1. If f_n is uniformly integrable, then f is integrable by Theorem 44 and Fatou's lemma. Define

$$f_{n,a} = 1_{\{|f_n| < a\}} f_n, \quad f_a = 1_{\{|f| < a\}} f.$$

It follows that $f_{n,a} \rightarrow f_a$, *a.e.* provided $\mu(|f| = a) = 0$. By bounded dominated convergence,

$$\int f_{n,a} d\mu \rightarrow \int f_a d\mu.$$

Writing

$$(6.11) \quad \int_{\{|f_n| \geq a\}} f_n d\mu = \int f_n d\mu - \int f_{n,a} d\mu$$

and

$$(6.12) \quad \int_{\{|f| \geq a\}} f d\mu = \int f d\mu - \int f_a d\mu,$$

we see that

$$\begin{aligned}
& \limsup_n \left| \int f_n d\mu - \int f d\mu \right| \\
& \leq \limsup_n \left| \int f_{n,a} d\mu - \int f_a d\mu \right| + \sup_n \int_{\{|f_n| \geq a\}} |f_n| d\mu + \int_{\{|f| \geq a\}} |f| d\mu \\
& = \sup_n \int_{\{|f_n| \geq a\}} |f_n| d\mu + \int_{\{|f| \geq a\}} |f| d\mu.
\end{aligned}$$

Note $\mu(|f| = a) = 0$ for all but countably many a . Sending $a \rightarrow \infty$ proves (6.10).

2. Suppose f_n, f are nonnegative integrable and (6.10) holds. Write

$$\int_{\{|f_n| \geq a\}} f_n d\mu = \int_{\{|f| \geq a\}} f d\mu + \left(\int_{\{|f_n| \geq a\}} f_n d\mu - \int_{\{|f| \geq a\}} f d\mu \right).$$

Since f is integrable, the first term is less than $\varepsilon/2$ when a is larger than some a_0 . If $\mu(|f| = a) = 0$, (6.11) and (6.12) indicate the term in the bracket is also less than $\varepsilon/2$ when n is larger than some n_0 .

Therefore,

$$\sup_{n > n_0} \int_{\{|f_n| \geq a\}} f_n d\mu \leq \varepsilon, \quad \forall a > a_0 \text{ with } \mu(|f| = a) = 0.$$

Since the finite family $\{f_1, \dots, f_{n_0}\}$ is uniformly integrable, the uniform integrability of $\{f_n, n \geq 1\}$ follows. \square

Additional details on the proof of Theorem 45. Suppose $|f_n(x)| \rightarrow |f(x)| < a$. Then for large n , $|f_n(x)| < a$. So $1_{\{|f_n| < a\}}$ and $1_{\{|f| < a\}}$ are both equal to 1, it follows $f_{n,a} \rightarrow f_a$ at x . The same is true for x with $|f(x)| > a$. If $|f(x)| = a \neq 0$, then $f_{n,a}(x) \rightarrow f_a(x)$ may not happen, since in this case $f_a(x) = 0$ while there could be a subsequence n_k with $f_{n_k}(x) < a$ so that

$$f_{n_k,a}(x) = f_{n_k}(x) \rightarrow f(x) \neq 0.$$

But if the set $\{x : |f(x)| = a\}$ has zero μ -measure, then $f_{n,a} \rightarrow f_a$, *a.e.* Fortunately the set of a for which $\mu(|f| = a)$ is not zero is at most countable. Indeed, let

$$F(x) = \mu(|f| \leq x).$$

Then $F(x)$ is non-decreasing, hence has at most countably many discontinuities. F is (right-continuous and thus) discontinuous at $x = a$ if and only if

$$\mu(|f| = a) = F(a) - F(a-) \neq 0.$$

This verifies that $\mu(|f| = a) = 0$ for all but countably many a .

COROLLARY 2. *Suppose that μ is finite, f_n, f are integrable. If $f_n \xrightarrow{a.e.} f$ or $f_n \xrightarrow{\mu} f$, then these are equivalent:*

(1) $\{f_n\}$ is uniformly integrable;

(2) $\int |f_n - f| d\mu \rightarrow 0$;

(3) $\int |f_n| d\mu \rightarrow \int |f| d\mu$.

6.5. Summary of various convergences.

$$f_n \xrightarrow{\mu} f \quad \begin{array}{c} \text{sub subseq} \\ \Longleftrightarrow \\ \text{Thm 20} \end{array} \quad f_n \xrightarrow{a.u.} f$$

Markov \Uparrow

$$f_n \xrightarrow{L_p} f \quad \begin{array}{c} \text{has a subseq} \\ \Longrightarrow \\ \Longleftarrow \\ \|f_n\|_p \rightarrow \|f\|_p \\ \text{Thm 40} \end{array} \quad f_n \xrightarrow{a.e.} f \quad \begin{array}{c} \xLeftrightarrow{\mu \text{ finite}} \\ \Longleftarrow \\ \text{Thm 19} \end{array} \quad f_n \xrightarrow{a.u.} f \quad \Downarrow \text{Thm 18}$$

$$f_n \xrightarrow{d} f \quad \begin{array}{c} \text{in a prob space} \\ \Longleftarrow \\ \text{Thm 24} \end{array} \quad f_n \xrightarrow{\mu} f$$

7. 概率空间的积分

7.1. Expected value. (Ω, \mathcal{F}, P) is a probability space, X a r.v.

Def 30. *Expectation, written EX ,*

$$EX = \int X dP.$$

Suppose X is discrete, i.e., X takes values in a finite or infinitely countable *distinct* sequence $\{x_1, x_2, \dots\}$. Then its expectation $(\int X dP$ computed according to (5.1)) equals

$$EX = \sum_i x_i P(X = x_i).$$

The mapping $i \mapsto P(X = x_i)$ is called the probability mass function of X . If $Y = g(X)$ for some measurable function g , then Y is discrete with values in, say, $\{y_1, y_2, \dots\}$. The expectation of Y , computed in the

same way as EX , is

$$EY = \sum_i y_i P(Y = y_i).$$

To calculate EY , we first need to find its probability mass function $i \mapsto P(Y = y_i)$. This can be complicated, and it is avoided by using the "*law of the unconscious statistician*",

$$EY = \sum_i g(x_i) P(X = x_i).$$

This turns out to be a change of variables formula (see also Theorem [48](#)).

Thm 46 (Change of variables formula). *Let (Ω, \mathcal{F}, P) be a probability space, X a r.v, and $g \in \mathcal{B}_{\mathbb{R}}$. If $g \geq 0$ or $\int_{\Omega} |g(X)| dP < \infty$, then*

$$(7.1) \quad Eg(X) = \int_{\Omega} g(X) dP = \int_{\mathbb{R}} g(x) d\mu_X.$$

Here $\mu_X(A) = PX^{-1}(A) = P(X \in A)$, $\forall A \in \mathcal{B}_{\mathbb{R}}$ is the probability induced by X (Section 5.3), which will be called the **distribution** of X .

PROOF. 1. The nonnegative case $g \geq 0$. If $g = 1_A$, then $g(X(\omega)) = 1_A(X(\omega)) = 1_{X^{-1}(A)}(\omega)$, so (7.1) reduces to the definition of μ_X . By linearity, (7.1) holds for simple functions. If g_n are simple functions such that $0 \leq g_n(x) \uparrow g(x)$, then $0 \leq g_n(X(\omega)) \uparrow g(X(\omega))$, then (7.1) follows by monotone convergence theorem.

2. The case $\int_{\Omega} |g(X)| dP < \infty$. Applying step 1 to $|g(X)|$ shows that g is integrable with respect to μ_X , hence the integrability of g^+ , g^- , and (7.1) follows from subtracting $Eg^-(X) = \int_{\mathbb{R}} g^-(x) d\mu_X$ from $Eg^+(X) = \int_{\mathbb{R}} g^+(x) d\mu_X$. □

The probability μ_X equals (as a result of the uniqueness Theorem 10) the measure μ constructed from the distribution function F

of $X : \mu((a, b]) = F(b) - F(a), \forall a, b$. The measure μ is called a Lebesgue-Stieltjes measure and its integral is the Lebesgue-Stieltjes integral (section 7.3). The above formula thus relates integral on a probability space to Lebesgue-Stieltjes integral over \mathbb{R} . The rightmost term of (7.1) is also written as $\int g dF$, i.e.

$$Eg(X) = \int_{\mathbb{R}} g(x) dF.$$

REMARK 2. *An implication of Theorem 46 is that the integration (e.g. the expectation and variance) of a random variable is a distributional property, i.e., it depends on the random variable only through its distribution. This lays the basis for applying probability theory tools such as Skorohod Theorem (Theorem 26).*

Def 31. *Variance, written $\text{Var}(X)$,*

$$\text{Var}(X) = \int (X - EX)^2 dP = E(X - EX)^2.$$

It is easy to see that

$$\text{Var}(X) = EX^2 - (EX)^2.$$

Def 32. *k-th moment, $k = 1, 2, \dots$,*

$$E(X^k) = \int X^k dP.$$

Example 14 (Bernoulli distribution). *Let $0 < p < 1$. $X \sim \text{Bernoulli}(p)$ if $P(X = 1) = p$, $P(X = 0) = 1 - p$. Then*

$$EX = 1 \cdot p + 0 \cdot (1 - p) = p.$$

$$\text{Var}(X) = EX^2 - (EX)^2 = p - p^2 = p(1 - p).$$

Example 15 (Poisson distribution). *Let $\lambda > 0$. $X \sim \text{Poisson}(\lambda)$ if*

$$P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}, \quad k = 0, 1, 2, \dots$$

Then

$$EX = \sum_{k=0}^{\infty} k \cdot e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^k}{(k-1)!} = \lambda.$$

$$E(X(X-1)) = \sum_{k=0}^{\infty} k(k-1) \cdot e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} \sum_{k=2}^{\infty} \frac{\lambda^k}{(k-2)!} = \lambda^2.$$

Hence $EX^2 = \lambda^2 + \lambda$, and

$$\text{Var}(X) = (\lambda^2 + \lambda) - \lambda^2 = \lambda.$$

Example 16 (Geometric distribution). *Repeatedly flip a coin with head probability p and stop only when the head appears. The number of tosses X has the distribution*

$$P(X = k) = (1 - p)^{k-1} p, \quad k = 1, 2, \dots$$

The distribution of X is called geometric, denoted by $X \sim \text{Geom}(p)$,

$$EX = \frac{1}{p}, \quad \text{Var}(X) = \frac{1}{p^2}.$$

7.2. Properties of expectation. X, Y are random variables. The following are immediate from section 6.1.

Jensen inequality: if X integrable, φ convex, then

$$\varphi(EX) \leq E\varphi(X).$$

Hölder inequality: if $p, q \geq 1, 1/p + 1/q = 1$, then

$$E|XY| \leq \|X\|_p \|Y\|_q.$$

Minkowski inequality: if $p \geq 1$, then

$$\|X + Y\|_p \leq \|X\|_p + \|Y\|_p.$$

Thm 47. $0 < s < t < \infty$, X is a r.v. Then $\|X\|_s \leq \|X\|_t$.

PROOF. By Hölder inequality with $p = \frac{t}{s}$, $q = \frac{t}{t-s}$,

$$\|X\|_s^s = E|X|^s \leq (E|X|^{sp})^{1/p} (E1^q)^{1/q} = (E|X|^t)^{s/t} = \|X\|_t^s.$$

□

Example 17. If X has $EX^2 < \infty$, then its expectation and variance exist, since $E|X| \leq \|X\|_2 < \infty$, and

$$0 \leq \text{Var}(X) \leq EX^2.$$

7.3. Lebesgue-Stieltjes and Riemann-Stieltjes integrals. Let G be a **generalized distribution function**, i.e., nondecreasing, right-continuous on \mathbb{R} . There is a unique measure μ such that

$$(7.2) \quad \mu((a, b]) = G(b) - G(a), \quad \forall a, b.$$

The measure μ constructed this way is called a **Lebesgue-Stieltjes measure**. Integration with respect to Lebesgue-Stieltjes measure is called **Lebesgue-Stieltjes integral**, denoted by $\int f d\mu$ or $\int f dG$.

REMARK 3. *Under suitable conditions (see below), $\int f dG$ may be interpreted as Riemann-Stieltjes integral. Since this does not provide anything new in the context of general measure theory, $\int f dG$ is best understood as a notional variant of $\int f d\mu$, and hence by convention (see (7.2)) $\int_a^b f dG$ means $\int_{(a,b]} f dG$ or $\int_{(a,b]} f d\mu$.*

Here we recall a few facts about Riemann-Stieltjes integration. Let G be the function as in (7.2), f a bounded function on $[a, b]$. Corresponding to each partition $\mathcal{P} : a = x_0 < x_1 < \cdots < x_n = b$, we consider

$$L(\mathcal{P}, f) = \sum_{i=1}^n \inf_{x \in [x_{i-1}, x_i]} f(x) \Delta G_i, \quad U(\mathcal{P}, f) = \sum_{i=1}^n \sup_{x \in [x_{i-1}, x_i]} f(x) \Delta G_i.$$

Here $\Delta G_i = G(x_i) - G(x_{i-1})$. Define

$$R_* f = \sup_{\mathcal{P}} L(\mathcal{P}, f), \quad R^* f = \inf_{\mathcal{P}} U(\mathcal{P}, f).$$

If $R_* f = R^* f$, then f is Riemann-Stieltjes integrable with respect to G , the common value, written $(R-S) \int f$, is called the Riemann-Stieltjes integral. For simplicity we have omitted the dependence of the integral on G in the notations.

A sufficient condition for Riemann-Stieltjes integrability is this: Suppose f is bounded on $[a, b]$, has at most finitely many discontinuities, G is continuous at every point where f is discontinuous. Then f is Riemann-Stieltjes integrable with respect to G .

Example 18. *If $a < s < b$, f is bounded on $[a, b]$, continuous at s and $G(x) = 1_{[s, \infty)}(x)$. Then*

$$(R-S) \int_a^b f dG = f(s).$$

Indeed, consider partitions $\mathcal{P} = \{x_0, x_1, x_2, x_3\}$, $a = x_0$ and $x_1 < x_2 = s < x_3 = b$. Then $\Delta G_2 = 1$, $\Delta G_i = 0$ if $i \neq 2$,

$$L(\mathcal{P}, f) = \inf_{x \in [x_1, x_2]} f(x), \quad U(\mathcal{P}, f) = \sup_{x \in [x_1, x_2]} f(x).$$

Since f is continuous at s , we see that $L(\mathcal{P}, f)$ and $U(\mathcal{P}, f)$ converge to $f(s)$ as $x_1 \rightarrow s$.

Thm 48. Suppose $c_n \geq 0$, $\sum c_n < \infty$, $\{s_n\}$ is a sequence of distinct points in (a, b) , and

$$G(x) = \sum_{n=1}^{\infty} c_n 1_{[s_n, \infty)}(x).$$

If f is continuous on $[a, b]$, then

$$(R-S) \int_a^b f dG = \sum_{n=1}^{\infty} c_n f(s_n).$$

PROOF. Exercise. □

If we denote by $L_* f$ the integral in (5.1) with the G -induced Lebesgue-Stieltjes measure in the role of μ , and by $L^* f$ the integral in (5.2). Then

$$R_* f \leq L_* f \leq L^* f \leq R^* f.$$

Therefore if, for instance, f is continuous on $[a, b]$, then it is Riemann-Stieltjes integrable, hence Lebesgue-Stieltjes integrable.

7.4. L_p convergence and uniform integrability.

Thm 49. (Ω, \mathcal{F}, P) is a probability space, $0 < p < \infty$, $X_n \in L_p$, $X \in \mathcal{F}$. If $X_n \xrightarrow{P} X$, then these are equivalent:

- (1) $\{|X_n|^p\}$ is uniformly integrable;
- (2) $X \in L_p$, $E(|X_n - X|^p) \rightarrow 0$;
- (3) $X \in L_p$, $E(|X_n|^p) \rightarrow E(|X|^p)$.

PROOF. 1. Suppose that $\{|X_n|^p\}$ is uniformly integrable. Observe that $X \in L_p$ by Theorem 45, hence $\{|X_n - X|^p\}$ is uniformly integrable since $|X_n - X|^p \leq C_p(|X_n|^p + |X|^p)$ where $C_p = 2^{p-1} \vee 1$. Note also that $|X_n - X|^p \xrightarrow{P} 0$. Therefore (1) implies (2) is a consequence of Theorem 45 with $f_n = |X_n - X|^p$.

2. (2) implies (3) because $\left| \|X_n\|_p - \|X\|_p \right| \leq \|X_n - X\|_p$, $0 < p < \infty$ (Theorem 36, Theorem 38).

3. (3) implies (2) follows from an application of Theorem 45 with $f_n = |X_n|^p$. □

We notice another criterion for uniform integrability, in addition to Theorem 44.

LEMMA 50. *Let (Ω, \mathcal{F}, P) be a probability space,*

$$\mathcal{H} = \{X_t : t \in T, E|X_t| < \infty\}.$$

Suppose that $g \geq 0$ is an increasing function on $[0, \infty)$ such that

$$\lim_{s \rightarrow \infty} \frac{g(s)}{s} = \infty$$

and

$$\sup_{X \in \mathcal{H}} \int g(|X|) dP < \infty.$$

Then \mathcal{H} is uniformly integrable.

PROOF. $\forall \varepsilon > 0$. Fix $a > 0$ so that

$$\frac{1}{a} \sup_{X \in \mathcal{H}} \int g(|X|) dP < \varepsilon.$$

There is $s_0 > 0$ such that $g(s) \geq as$ for all $s \geq s_0$. Hence, $\forall X \in \mathcal{H}$, $s \geq s_0$,

$$\int_{\{|X| \geq s\}} |X| dP \leq \frac{1}{a} \int_{\{|X| \geq s\}} g(|X|) dP \leq \frac{1}{a} \sup_{X \in \mathcal{H}} \int g(|X|) dP < \varepsilon.$$

□

8. 乘积测度空间

Let (X, \mathcal{X}, μ) , (Y, \mathcal{Y}, ν) be measure spaces. The problem is to construct a **product measure** π on $X \times Y$ such that

$$(8.1) \quad \pi(A \times B) = \mu(A)\nu(B) \text{ for } A \in \mathcal{X}, B \in \mathcal{Y}.$$

8.1. Product σ -field.

Def 33. *In the product space $X \times Y$, a **measurable rectangle** is a product of the form*

$$A \times B, A \in \mathcal{X}, B \in \mathcal{Y}.$$

Let

$$(8.2) \quad \mathcal{S} = \{A \times B : A \in \mathcal{X}, B \in \mathcal{Y}\}$$

*be the class of measurable rectangles on $X \times Y$. The **product σ -field** on $X \times Y$ is then defined as*

$$\mathcal{X} \times \mathcal{Y} = \sigma(\mathcal{S}).$$

The space $X \times Y$ equipped with this product σ -field is called **product measurable space**.

As the example below shows, the product σ -field is generally larger than the class of measurable rectangles.

Example 19. $\mathcal{B}(\mathbb{R}^2) = \mathcal{B}(\mathbb{R}) \times \mathcal{B}(\mathbb{R})$. Here $\mathcal{B}(\mathbb{R}^2)$ is the usual σ -field on \mathbb{R}^2 generated by the class of products of one-dimensional intervals.

Example 20. If $A \times B \in \mathcal{S}$, then

$$(A \times B)^c = A^c \times Y + A \times B^c \in \mathcal{S}.$$

From this it is easy to check that \mathcal{S} is a semi-ring and $X \times Y \in \mathcal{S}$.

Def 34. The **section of a set** $E \in X \times Y$ at $x \in X$ is

$$E_x = \{y : (x, y) \in E\}.$$

Similarly $E_y = \{x : (x, y) \in E\}$ is the section at $y \in Y$. The **section of a function** $f(x, y)$ at $x \in X$ is the mapping

$$y \mapsto f(x, y).$$

The section at $y \in Y$ is $x \mapsto f(x, y)$.

Example 21. If $E, E_k \in \mathcal{X} \times \mathcal{Y}$, $x \in X$, then

$$(E^c)_x = (E_x)^c, \quad \left(\bigcup_k E_k\right)_x = \bigcup_k (E_k)_x, \quad \left(\bigcap_k E_k\right)_x = \bigcap_k (E_k)_x$$

Thm 51. (1) Sections of $\mathcal{X} \times \mathcal{Y}$ -measurable set are measurable.

(2) Sections of $\mathcal{X} \times \mathcal{Y}$ -measurable function are measurable.

PROOF. 1. Fix $x \in X$. Consider the mapping $R_x : Y \mapsto X \times Y$ defined by $R_x(y) = (x, y)$. We intend to prove that R_x is \mathcal{Y} -measurable so that the conclusion follows immediately:

$$E_x = R_x^{-1}E \in \mathcal{Y}, \quad \forall E \in \mathcal{X} \times \mathcal{Y}.$$

If $E = A \times B$ is a measurable rectangle, then $R_x^{-1}E = B \in \mathcal{Y}$. This shows that R_x is \mathcal{Y} -measurable, since $\mathcal{X} \times \mathcal{Y}$ is generated by measurable rectangles. So the first part is proved.

2. If $f \in \mathcal{X} \times \mathcal{Y}$, then $f(x, \cdot) = f \circ R_x(\cdot)$ is \mathcal{Y} -measurable by measurable composition.

3. The conclusion for fixed $y \in Y$ is proved similarly. □

8.2. Product measure space. Let $(X, \mathcal{X}, \mu), (Y, \mathcal{Y}, \nu)$ be measure spaces.

LEMMA 52. *Suppose that μ and ν are finite. If $E \in \mathcal{X} \times \mathcal{Y}$, then the mapping $x \mapsto \nu(E_x)$ is \mathcal{X} -measurable, $y \mapsto \mu(E_y)$ is \mathcal{Y} -measurable.*

PROOF. Let \mathcal{L} be the class of $E \in \mathcal{X} \times \mathcal{Y}$ that has the stated property. Then \mathcal{L} is a λ -system. Indeed, it is easy to see that $X \times Y \in \mathcal{L}$. If $E, F \in \mathcal{L}$ with $E \subset F$, then

$$x \mapsto \nu((F \setminus E)_x) = \nu(F_x \setminus E_x) = \nu(F_x) - \nu(E_x)$$

is \mathcal{X} -measurable (the finiteness of ν is used to justify subtraction). If $E_k \in \mathcal{L}$, $E_k \subset E_{k+1}$, then

$$x \mapsto \nu\left(\left(\bigcup_k E_k\right)_x\right) = \nu\left(\bigcup_k (E_k)_x\right) = \lim_k \nu((E_k)_x)$$

is \mathcal{X} -measurable. This shows that \mathcal{L} is a λ -system. For any measurable rectangle $E = A \times B$, the function

$$x \mapsto \nu(E_x) = 1_A(x)\nu(B)$$

is \mathcal{X} -measurable. So \mathcal{L} contains the π -system of measurable rectangles, thus coincides with $\mathcal{X} \times \mathcal{Y}$ by π - λ Theorem. \square

LEMMA 53. *Let (X, \mathcal{X}, μ) , (Y, \mathcal{Y}, ν) be σ -finite measure spaces. Define*

$$\pi_{21}(E) = \int_X \nu(E_x) d\mu, \quad E \in \mathcal{X} \times \mathcal{Y}$$

and

$$\pi_{12}(E) = \int_Y \mu(E_y) d\nu, \quad E \in \mathcal{X} \times \mathcal{Y}.$$

Then $\pi_{21}(E) = \pi_{12}(E)$ for $E \in \mathcal{X} \times \mathcal{Y}$. Moreover, π_{21} , π_{12} satisfy (8.1).

PROOF. 1. First suppose μ, ν are finite. Let \mathcal{L} be the class of $E \in \mathcal{X} \times \mathcal{Y}$ such that $\pi_{21}(E) = \pi_{12}(E)$. \mathcal{L} contains all measurable

rectangle $A \times B$, since

$$\begin{aligned}\pi_{21}(A \times B) &= \int_X 1_A(x) \nu(B) d\mu = \mu(A) \nu(B) \\ &= \int_Y \mu(A) 1_B(y) d\nu = \pi_{12}(A \times B).\end{aligned}$$

It is easy to check that \mathcal{L} is a λ -system, hence equals $\mathcal{X} \times \mathcal{Y}$ by π - λ Theorem.

2. Now suppose μ, ν are σ -finite, then there are $\{A_m\}, \{B_n\}$ that partition X and Y into disjoint sets of finite measure. Define

$$\mu_m(E) = \mu(E \cap A_m), \quad \nu_n(E) = \nu(E \cap B_n), \quad E \in \mathcal{X} \times \mathcal{Y}.$$

Step 1 is valid for these finite measures,

$$(8.3) \quad \pi_{21}^{(mn)}(E) = \pi_{12}^{(mn)}(E), \quad E \in \mathcal{X} \times \mathcal{Y},$$

where

$$\pi_{21}^{(mn)}(E) = \int_X \nu_n(E_x) d\mu_m,$$

and

$$\pi_{12}^{(mn)}(E) = \int_Y \mu_m(E_y) d\nu_n.$$

In addition, for measurable rectangle $A \times B$,

$$(8.4) \quad \pi_{21}^{(mn)}(A \times B) = \mu_m(A) \nu_n(B) = \pi_{12}^{(mn)}(A \times B).$$

From Lemma 52, $x \mapsto \nu_n(E_x)$ is measurable. Since $\nu = \sum_n \nu_n$, $x \mapsto \nu(E_x)$ is measurable. The same can be said for $y \mapsto \mu(E_y)$. Therefore π_{21} , π_{12} are well-defined for the σ -finite case. By Example 10, 11 and (8.3), for $E \in \mathcal{X} \times \mathcal{Y}$,

$$\begin{aligned} \pi_{21}(E) &= \int_X \nu(E_x) d\mu = \sum_m \int_X \nu(E_x) d\mu_m = \sum_m \int_X \sum_n \nu_n(E_x) d\mu_m \\ &= \sum_{m,n} \pi_{21}^{(mn)}(E) = \sum_{m,n} \pi_{12}^{(mn)}(E) = \sum_n \int_Y \sum_m \mu_m(E_y) d\nu_n \\ &= \pi_{12}(E). \end{aligned}$$

Particularly, this together with (8.4) yields that, for measurable rectangle $A \times B$,

$$\pi_{21}(A \times B) = \sum_{m,n} \mu_m(A) \nu_n(B) = \pi_{12}(A \times B).$$

□

Thm 54. *Let (X, \mathcal{X}, μ) , (Y, \mathcal{Y}, ν) be σ -finite measure spaces. Then*

$$\pi(E) \triangleq \pi_{21}(E) = \pi_{12}(E)$$

defines the unique σ -finite measure on $X \times Y$ that satisfies (8.1).

PROOF. Decompose X and Y into disjoint sets $\{A_m\}$, $\{B_n\}$ of finite measure, and define μ_m , ν_n as before. Then $X \times Y$ is the disjoint union of $\{A_m \times B_n\}$ and each $A_m \times B_n$ has finite π -measure: $\pi(A_m \times B_n) = \mu_m(A_m) \nu_n(B_n)$. It follows that π is σ -finite. The uniqueness is a consequence of Theorem 10, since measurable rectangles form a π -system (Example 20). □

In the future, the product measure π of μ and ν will be written as $\mu \times \nu$.

8.3. Fubini's Theorem.

Thm 55. *Let (X, \mathcal{X}, μ) , (Y, \mathcal{Y}, ν) be σ -finite measure spaces, π the product measure constructed in Theorem 54, f a $\mathcal{X} \times \mathcal{Y}$ -measurable function. If f is nonnegative or $\int_{X \times Y} |f| d\pi < \infty$. Then*

$$\int_X \left[\int_Y f(x, y) d\nu \right] d\mu = \int_Y \left[\int_X f(x, y) d\mu \right] d\nu = \int_{X \times Y} f d\pi.$$

It is implicit in the statement that all integrands are integrable. In the nonnegative case, if one of the above integrals is infinite, so it is with the other two.

PROOF. The conclusion holds for measurable indicator function by Theorem 54, and hence simple function by linearity of integration.

Then the monotone convergence theorem gives the conclusion for non-negative measurable function. If $\int_{X \times Y} |f| d\pi < \infty$, then applying the nonnegative case to $|f|$,

$$\int_X \left[\int_Y |f| d\nu \right] d\mu = \int_{X \times Y} |f| d\pi < \infty.$$

It follows that

$$\int_Y |f| d\nu < \infty, \text{ a.e. } x.$$

Hence it makes sense (outside a set of zero μ -measure) to write

$$\int_Y f d\nu = \int_Y f^+ d\nu - \int_Y f^- d\nu.$$

Now the desired property follows by integrating over X and using the result for nonnegative integrand. The same reasoning applies to

$$\int_Y \left[\int_X |f| d\mu \right] d\nu.$$

□

8.4. Applications.

Example 22 (Euler-Poisson integral). Let $I = \int_{\mathbb{R}} e^{-x^2} dx$. By Fubini's theorem

$$I^2 = \iint_{\mathbb{R}^2} e^{-(x+y)^2} dx dy = \iint_{\substack{r \geq 0, \\ 0 \leq \theta < 2\pi}} e^{-r^2} r dr d\theta.$$

Again by Fubini's theorem, the double integral on the RHS is written as an iterated integral and evaluated to give $I^2 = \pi$, so

$$I = \int_{\mathbb{R}} e^{-x^2} dx = \sqrt{\pi}.$$

Thm 56. Let (X, \mathcal{X}, μ) be a σ -finite measure space, $f \geq 0$ measurable. Then

$$\int f d\mu = \int_0^\infty \mu(f \geq t) dt = \int_0^\infty \mu(f > t) dt$$

PROOF. Since f is nonnegative, we may write (recall our convention Remark 3), $\forall x$,

$$f(x) = \int_0^{f(x)} dt = \int_0^\infty 1_{(0, f(x)]}(t) dt.$$

Notice that

$$1_{(0, f(x)]}(t) = 1_{\{x: f(x) \geq t\}}(x).$$

Then using Fubini theorem

$$\begin{aligned} \int f d\mu &= \int \int_0^\infty 1_{(0, f(x)]}(t) dt d\mu = \int_0^\infty \int 1_{(0, f(x)]}(t) d\mu dt \\ &= \int_0^\infty \int 1_{\{x: f(x) \geq t\}}(x) d\mu dt = \int_0^\infty \mu(\{x : f(x) \geq t\}) dt. \end{aligned}$$

Since the set of t such that $\mu(\{x : f(x) = t\})$ is non-zero is at most countable, hence has zero Lebesgue measure. Thus the two integrals are equal,

$$\int_0^\infty \mu(\{x : f(x) \geq t\}) dt = \int_0^\infty \mu(\{x : f(x) > t\}) dt.$$

□

If f takes values in $\{y_1, y_2, \dots\}$ and $0 \leq y_1 < y_2 < \dots$. Then $t \mapsto \mu(f \geq t)$ is a step function

$$\mu(f \geq t) = \begin{cases} \mu(f \geq y_1), & 0 \leq t \leq y_1; \\ \mu(f \geq y_n), & y_{n-1} < t \leq y_n. \end{cases}$$

Hence Theorem 56 reduces to

$$\int f d\mu = y_1 \mu(f \geq y_1) + \sum_{n=2}^{\infty} (y_n - y_{n-1}) \mu(f \geq y_n).$$

A particular case of this is f taking values in nonnegative integers.

COROLLARY 3. *Let (X, \mathcal{X}, μ) be a σ -finite measure space, f measurable with values in $\{0, 1, 2, \dots\}$. Then*

$$\int f d\mu = \sum_{n=1}^{\infty} \mu(f \geq n) = \sum_{n=0}^{\infty} \mu(f > n).$$

Example 23. *Example 11 has validated interchanging the order of summation as an application of dominated convergence theorem. The same can also be proved directly using Fubini theorem.*

Thm 57 (Integration by parts). *Let F, G be two nondecreasing, right-continuous functions on \mathbb{R} , then, for $a < b$,*

$$F(b)G(b) - F(a)G(a) = \int_{(a,b]} G(x)dF(x) + \int_{(a,b]} F(x-)dG(x),$$

or equivalently

$$\begin{aligned} F(b)G(b) - F(a)G(a) &= \int_{(a,b]} G(-x)dF(x) + \int_{(a,b]} F(x-)dG(x) \\ &\quad + \sum_{a < x \leq b} \Delta F(x)\Delta G(x), \end{aligned}$$

where $\Delta F(x) = F(x) - F(x-)$.

PROOF. Denote respectively by μ, ν the Lebesgue-Stieltjes measure induced by F, G . Let $\pi = \mu \times \nu$ be the product measure of μ, ν .

Using Fubini Theorem we have

$$\begin{aligned}
 \pi((x, y) : a < x < y \leq b) &= \int_{(a,b]} \int_{(a,y)} d\mu(x) d\nu(y) \\
 &= \int_{(a,b]} [F(y-) - F(a)] d\nu(y) \\
 &= \int_{(a,b]} F(y-) d\nu(y) - F(a)[G(b) - G(a)]
 \end{aligned}$$

and similarly

$$\begin{aligned}
 \pi((x, y) : a < y \leq x \leq b) &= \int_{(a,b]} \int_{(a,x]} d\nu(y) d\mu(x) \\
 &= \int_{(a,b]} [G(x) - G(a)] d\mu(x) \\
 &= \int_{(a,b]} G(x) d\mu(x) - G(a)[F(b) - F(a)]
 \end{aligned}$$

By the construction of π ,

$$(F(b) - F(a))(G(b) - G(a)) = \pi((x, y) \in (a, b] \times (a, b]).$$

The first conclusion follows by putting together these equations. To complete the proof, it suffices to note that

$$\begin{aligned} \pi((x, y) : a < y = x \leq b) &= \int_{(a, b]} \nu(\{x\}) d\mu(x) \\ &= \int_{(a, b]} [G(x) - G(x-)] d\mu(x) \\ &= \sum_{a < x \leq b} \Delta F(x) \Delta G(x). \end{aligned}$$

□

8.5. Finite-dimensional product space. The discussion for two-dimensional product space obviously extends to finite dimensions. Denote by $\mathcal{B}(\mathbb{R}^n)$ the σ -field on \mathbb{R}^n generated by open sets. The product

σ -field $\prod_{i=1}^n \mathcal{B}(\mathbb{R})$ is defined as being generated by measurable rectangles with sides in $\mathcal{B}(\mathbb{R})$. Similar to Example [19](#), we have on \mathbb{R}^n ,

$$(8.5) \quad \mathcal{B}(\mathbb{R}^n) = \prod_{i=1}^n \mathcal{B}(\mathbb{R}).$$

9. 独立性 Independence

Let (Ω, \mathcal{F}, P) be a probability space. A subset in \mathcal{F} is called an **event**, and an element of Ω is called a **sample**.

9.1. Independence of events and random variables.

Def 35. *The events A and B are independent if*

$$P(A \cap B) = P(A)P(B).$$

Thm 58. *A, B are independent if and only if one of the three pairs are independent: (i) A^c, B ; (ii) A, B^c ; (iii) A^c, B^c .*

Example 24. *An event A is independent of any event if and only if $P(A) = 0$ or 1 . In particular, Ω and \emptyset are independent of any event.*

Example 25 (Gambler's fallacy). *In some situations, an individual erroneously think that certain event is more or less likely to happen in the future based on the outcome of the past events. This incorrect belief may lead a gambler in a coin flipping game to believe that after 100 successive heads, the next toss would be more likely to*

come up tail. The fallacy roots from the ignorance of the independence between tosses.

Def 36. σ -fields \mathcal{F}, \mathcal{G} are independent if

$$P(A \cap B) = P(A)P(B), \quad \forall A \in \mathcal{F}, B \in \mathcal{G}.$$

Def 37. Random variables X, Y are independent if $\sigma(X)$ and $\sigma(Y)$ are independent, where $\sigma(X) = X^{-1}(\mathcal{B}(\mathbb{R}))$.

Thm 59. A, B are independent if and only if 1_A and 1_B are independent.

PROOF. Note that $\sigma(1_A) = \{A, A^c, \emptyset, \Omega\}$ and the same for $\sigma(1_B)$. □

Def 38. A family of events $\{A_1, \dots, A_n\}$ is independent if for any $I \subset \{1, \dots, n\}$,

$$P\left(\bigcap_{i \in I} A_i\right) = \prod_{i \in I} P(A_i).$$

Pairwise independence is weaker than independence.

Example 26. *Flip a fair coin twice and consider the events,*

$$A_1 = \{ \text{head-head, head-tail} \},$$

$$A_2 = \{ \text{head-head, tail-head} \},$$

$$A_3 = \{ \text{head-head, tail-tail} \}.$$

Then A_1, A_2, A_3 are pairwise independent but not independent, since

$$P(A_i \cap A_j) = P(\text{head-head}) = \frac{1}{4} = P(A_i)P(A_j), \quad i \neq j.$$

$$P(A_1 \cap A_2 \cap A_3) = P(\text{head-head}) = \frac{1}{4} \neq \frac{1}{8} = P(A_1)P(A_2)P(A_3).$$

Def 39. $\mathcal{A}_1, \dots, \mathcal{A}_n$ are classes of sets. They are independent if for any $I \subset \{1, \dots, n\}$,

$$P\left(\bigcap_{i \in I} A_i\right) = \prod_{i \in I} P(A_i), \quad \forall A_i \in \mathcal{A}_i.$$

If we denote by \mathcal{A}'_i the class formed by augmenting \mathcal{A}_i with Ω . Then it is easy to see that $\mathcal{A}_1, \dots, \mathcal{A}_n$ are independent if and only if

$\mathcal{A}'_1, \dots, \mathcal{A}'_n$ are independent. Definition 39 is thus equivalent to the *full-product form*,

$$(9.1) \quad P\left(\bigcap_{i=1}^n A_i\right) = \prod_{i=1}^n P(A_i), \quad \forall A_i \in \mathcal{A}'_i.$$

This form may bring added convenience when independence is to be verified. Since Ω is contained in σ -field, the independence of random variables can be defined in this full-product form.

Def 40. X_1, \dots, X_n are independent if $\sigma(X_1), \dots, \sigma(X_n)$ are independent, i.e.,

$$P\left(\bigcap_{i=1}^n \{X_i \in A_i\}\right) = \prod_{i=1}^n P(X_i \in A_i), \quad \forall A_i \in \mathcal{B}(\mathbb{R}), \quad i = 1, \dots, n.$$

Thm 60. Suppose that $\mathcal{A}_1, \mathcal{A}_2, \dots, \mathcal{A}_n$ are independent π -systems. Then $\sigma(\mathcal{A}_1), \sigma(\mathcal{A}_2), \dots, \sigma(\mathcal{A}_n)$ are independent.

PROOF. 1. Clearly it suffices to show that $\sigma(\mathcal{A}_1), \mathcal{A}_2, \dots, \mathcal{A}_n$ are independent, since the conclusion applies to itself and would yield that $\sigma(\mathcal{A}_1), \sigma(\mathcal{A}_2), \dots, \mathcal{A}_n$ are independent, and so on.

2. Now we show that $\sigma(\mathcal{A}_1), \mathcal{A}_2, \dots, \mathcal{A}_n$ are independent. Fix $A_i \in \mathcal{A}_i, i = 2, \dots, n$. Let $E = \bigcap_{i=2}^n A_i$ and

$$\mathcal{L}_E = \{A \in \sigma(\mathcal{A}_1) : P(A \cap E) = P(A)P(E)\}.$$

Then $\mathcal{A}_1 \subset \mathcal{L}_E$. In view of Example 24, $\Omega \in \mathcal{L}_E$. If $B_1, B_2 \in \mathcal{L}_E$ and $B_1 \subset B_2$, then

$$\begin{aligned} P((B_2 - B_1) \cap E) &= P(B_2 \cap E) - P(B_1 \cap E) \\ &= P(B_2)P(E) - P(B_1)P(E) \\ &= P(B_2 - B_1)P(E). \end{aligned}$$

Hence $B_2 - B_1 \in \mathcal{L}_E$. Finally let $B_k \in \mathcal{L}_E$, $B_k \subset B_{k+1}$, then

$$\begin{aligned} P\left(\left(\bigcup_{k=1}^{\infty} B_k\right) \cap E\right) &= \lim_k P(B_k \cap E) \\ &= \lim_k P(B_k)P(E) = P\left(\bigcup_{k=1}^{\infty} B_k\right)P(E). \end{aligned}$$

Thus $\bigcup_{k=1}^{\infty} B_k \in \mathcal{L}_E$. Therefore \mathcal{L}_E is a λ -system and $\sigma(\mathcal{A}_1) \subset \mathcal{L}_E$. The desired conclusion follows. \square

Since $\mathcal{B}(\mathbb{R})$ is generated by the class $\mathcal{S} = \{(-\infty, a] : a \in \mathbb{R}\}$, the σ -field $\sigma(X)$ generated by X equals $\sigma(\{X \leq a : a \in \mathbb{R}\})$, which together with Theorem 60 gives the following criterion for independence in terms of distribution functions.

Thm 61. X_1, \dots, X_n are independent if and only if $\forall x_1, \dots, x_n \in \mathbb{R}$,

$$P(X_1 \leq x_1, \dots, X_n \leq x_n) = \prod_{i=1}^n P(X_i \leq x_i).$$

Note the class $\{X_i \leq x_i : x_i \in \mathbb{R}\}$ may not contain but can approximate Ω , so it is still legal to use the full-product form (9.1).

Recall from Theorem 46 that each random variable X induces a probability μ on \mathbb{R} , which is called the distribution of X . When random vector (X_1, \dots, X_n) is involved, the same can be said. From Section 5.3, we see that the random vector induces a probability on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$, where $\mathcal{B}(\mathbb{R}^n)$ equals the σ -field generated by measurable rectangles with sides in $\mathcal{B}(\mathbb{R})$ (see (8.5)). The induced probability on $\mathcal{B}(\mathbb{R}^n)$, denoted by P_{X_1, \dots, X_n} , satisfies, $\forall B_1, \dots, B_n \in \mathcal{B}(\mathbb{R})$,

$$P_{X_1, \dots, X_n}(B_1 \times \cdots \times B_n) = P(X_1 \in B_1, \dots, X_n \in B_n).$$

P_{X_1, \dots, X_n} is called the **joint distribution** of (X_1, \dots, X_n) . If B_i takes the form $(-\infty, x_i]$, then we get an associated mapping

$$F_{X_1, \dots, X_n} : (x_1, \dots, x_n) \mapsto P(X_1 \leq x_1, \dots, X_n \leq x_n)$$

which is called the **joint distribution function**² of the random vectors (X_1, \dots, X_n) . Moreover, the general change of variables formula from Section 5.3 tells us that, for measurable $g : \mathbb{R}^n \mapsto \mathbb{R}$,

$$(9.2) \quad Eg(X_1, \dots, X_n) = \int_{\mathbb{R}^n} g(x_1, \dots, x_n) dP_{X_1, \dots, X_n}.$$

whenever one of the integrals exists.

Example 27. *The discrete random variables $X \in \{x_1, x_2, \dots\}$ and $Y \in \{y_1, y_2, \dots\}$ are independent if and only if*

$$(9.3) \quad P(X = x_i, Y = y_j) = P(X = x_i)P(Y = y_j), \quad \forall i, j.$$

²In fact, by uniqueness P_{X_1, \dots, X_n} equals the Lebesgue-Stieltjes measure determined by the joint distribution function on $\mathcal{B}(\mathbb{R}^n)$.

PROOF. 1. Suppose that (9.3) holds. $\forall x, y \in \mathbb{R}$,

$$\begin{aligned} P(X \leq x, Y \leq y) &= \sum_{\substack{i: x_i \leq x \\ j: y_j \leq y}} P(X = x_i, Y = y_j) \\ &= \sum_{\substack{i: x_i \leq x \\ j: y_j \leq y}} P(X = x_i)P(Y = y_j). \end{aligned}$$

The double summation may contain infinite number of terms, but we can invoke Fubini theorem to write it as iterated summation (see also Example 23)

$$\begin{aligned} \sum_{\substack{i: x_i \leq x \\ j: y_j \leq y}} P(X = x_i)P(Y = y_j) &= \sum_{i: x_i \leq x} P(X = x_i) \sum_{j: y_j \leq y} P(Y = y_j) \\ &= P(X \leq x)P(Y \leq y). \end{aligned}$$

Hence

$$P(X \leq x, Y \leq y) = P(X \leq x)P(Y \leq y), \quad \forall x, y \in \mathbb{R}.$$

By Theorem 61 X and Y are independent.

2. Conversely suppose X and Y are independent. Then $\forall i, j$, the events $\{X = x_i\}$, $\{Y = y_j\}$ are independent, so (9.3) holds. \square

In general, the probability of the event $\{X \in (a, b], Y \in (c, d]\}$ can be expressed in terms of joint distribution function F ,

$$P(X \in (a, b], Y \in (c, d]) = F(b, d) - F(b, c) - F(a, d) + F(a, c).$$

Thm 62. *Suppose that the collection of events*

$$A_{IJ} = \{A_{ij} : i \in I, j \in J\}$$

are independent. Here I, J are finite or infinite index sets. Let

$$\mathcal{F}_i = \sigma(A_{ij} : j \in J), \forall i \in I.$$

Then $\mathcal{F}_1, \mathcal{F}_2, \dots$ are independent.

REMARK 4. *Independence of infinite number of events is defined as any finite subcollection being independent.*

PROOF. $\forall i \in I$, denote

$$\mathcal{A}_i = \{\text{all finite intersections of } A_{i1}, A_{i2}, \dots\}.$$

Then $\mathcal{A}_1, \mathcal{A}_2, \dots$ are π -systems and $\mathcal{F}_i = \sigma(\mathcal{A}_i)$. Hence the conclusion follows from Theorem 60. \square

By inspecting the proof, we see that the above theorem extends to the case where the index set J varies with $i \in I$. As an application, we have the following useful result which states that functions of disjoint subgroups of independent random variables are independent.

For ease of writing, we introduce the notation of indexing by set, for example,

if $I = \{i_1, \dots, i_l\}$, then X_I means $(X_{i_1}, \dots, X_{i_l})$.

Thm 63. *Divide the independent random variables X_1, \dots, X_n into disjoint subgroups X_{I_1}, \dots, X_{I_k} , where $I_1, I_2, \dots, I_k \subset \{1, 2, \dots, n\}$ are disjoint, $\bigcup_{i=1}^k I_i = \{1, 2, \dots, n\}$. If $g_1(x_{I_1}), \dots, g_k(x_{I_k})$ are measurable functions, then $g_1(X_{I_1}), \dots, g_k(X_{I_k})$ are independent.*

PROOF. Note $\sigma(g_s(X_{I_s})) \subset \sigma(X_{I_s})$, $s \in \{1, \dots, k\}$, hence it is enough to show that $\sigma(X_{I_1}), \dots, \sigma(X_{I_k})$ are independent. Each $\sigma(X_{I_i})$ can be generated by $\{\sigma(X_j) : j \in I_i\}$. Now the proof is completed by applying Theorem 62 to

$$A_{IJ} = \{\sigma(X_j) : j \in I_i\}.$$

□

9.2. Independence and expectation.

Thm 64. *Suppose that X_1, \dots, X_n are independent with respective distribution μ_i . Then (X_1, \dots, X_n) has the joint distribution $\mu_1 \times \dots \times \mu_n$.*

PROOF. $\forall B_1, \dots, B_n \in \mathcal{B}(\mathbb{R})$, using independence and the definition of product measure

$$\begin{aligned} P(X_1 \in B_1, \dots, X_n \in B_n) &= \prod_{i=1}^n P(X_i \in B_i) = \prod_{i=1}^n \mu_i(B_i) \\ &= \mu_1 \times \dots \times \mu_n(B_1 \times \dots \times B_n). \end{aligned}$$

The class of measurable rectangles

$$\{B_1 \times \cdots \times B_n : B_1, \dots, B_n \in \mathcal{B}(\mathbb{R})\}$$

is a π -system. Therefore by uniqueness (Theorem 10), $\mu_1 \times \cdots \times \mu_n$ agrees with the joint distribution of (X_1, \dots, X_n) . \square

Thm 65. *Suppose that X, Y are independent with respective distribution μ and ν , $h : \mathbb{R}^2 \mapsto \mathbb{R}$ is measurable. If $h \geq 0$ or $E|h(X, Y)| < \infty$, then*

$$Eh(X, Y) = \int_{\mathbb{R}^2} h(x, y) d\mu(x) d\nu(y).$$

In particular if $h(x, y) = f(x)g(y)$, then

$$Ef(X)g(Y) = Ef(X) \cdot Eg(Y).$$

PROOF. By Theorem 64, the induced probability of (X, Y) is given by the product of μ and ν . Then using the general change of variables formula (Section 5.3, Formula 9.2), we can write

$$Eh(X, Y) = \int_{\mathbb{R}^2} h(x, y) d\mu(x) d\nu(y)$$

The remaining conclusion follows easily. □

Inductively using the above theorem we get the following expectation formula for independent random variables.

Thm 66. *Suppose that X_1, \dots, X_n are independent and either (a) $X_i \geq 0, \forall i$ or (b) $E|X_i| < \infty, \forall i$. Then*

$$EX_1 \cdots X_n = EX_1 \cdots EX_n.$$

PROOF. The nonnegative case is immediate from Theorem 65. To prove case (b), applying the nonnegative case to $|X_1|$ and $|X_2|$, we have

$$E|X_1 X_2| = E|X_1| \cdot E|X_2|.$$

Since the RHS is finite, so $E|X_1 X_2| < \infty$ and by Theorem 65,

$$EX_1 X_2 = EX_1 \cdot EX_2.$$

Now the nonnegative case is again applicable to $|X_1 X_2|$ and $|X_3|$,

$$E|X_1 X_2 X_3| = E|X_1 X_2| \cdot E|X_3|.$$

So $E|X_1X_2X_3| < \infty$ and Theorem 65 can be invoked to get

$$EX_1X_2X_3 = EX_1 \cdot EX_2 \cdot EX_3.$$

The procedure continues until X_n is processed, then the proof is completed. \square

Def 41. *The covariance of X and Y is defined as*

$$\text{Cov}(X, Y) = E(X - EX)(Y - EY) = EXY - EX \cdot EY.$$

Def 42. *X and Y are uncorrelated if $\text{Cov}(X, Y) = 0$, i.e.,*

$$EXY = EX \cdot EY.$$

That X and Y are independent implies that they are uncorrelated, but not vice versa.

Example 28. *Suppose that X, Y are jointly distributed as below*

	$Y = -1$	$Y = 0$	$Y = 1$
$X = -1$	0	1/4	0
$X = 0$	1/4	0	1/4
$X = 1$	0	1/4	0

Then $EX = EY = EXY = 0$, but X, Y are not independent by Example 27, since

$$P(X = 0, Y = 0) = 0 \neq \frac{1}{4} = P(X = 0)P(Y = 0).$$

Thm 67. Suppose that X_1, \dots, X_n are pairwise uncorrelated and $EX_i^2 < \infty$. Then

$$\text{Var}(X_1 + \dots + X_n) = \text{Var}(X_1) + \text{Var}(X_2) + \dots + \text{Var}(X_n).$$

PROOF. Denote $X = X_1 + \dots + X_n$. We have

$$\begin{aligned} \text{Var}(X) &= E(X - EX)^2 = E\left(\sum_{i=1}^n (X_i - EX_i)\right)^2 \\ &= E\left(\sum_{i=1}^n (X_i - EX_i)^2 + \sum_{i \neq j} (X_i - EX_i)(X_j - EX_j)\right) \\ &= \text{Var}(X_1) + \text{Var}(X_2) + \dots + \text{Var}(X_n). \end{aligned}$$

We have used that X_1, \dots, X_n are pairwise uncorrelated, hence $\forall i \neq j$,

$$E((X_i - EX_i)(X_j - EX_j)) = EX_iX_j - EX_iEX_j = 0.$$

□

9.3. Sum of independent random variables. "Independent and identically distributed" is abbreviated as **i.i.d.**

Example 29. (*Binomial distribution*) Let $p \in (0, 1)$, X_1, \dots, X_n i.i.d. $\sim \text{Bernoulli}(p)$. Define $S_n = \sum_{i=1}^n X_i$. Then

$$P(S_n = k) = C_n^k p^k (1 - p)^{n-k}, \quad k = 0, \dots, n.$$

The distribution of S_n is the binomial distribution, written $S_n \sim \text{Bin}(n, p)$. By employing linearity of integration, Example 14 and Theorem 67,

$$ES_n = np, \quad \text{Var}(S_n) = np(1 - p).$$

Example 30 (The problem of points). A coin with head probability p is flipped repeatedly. Gambler A wins one point if head appears

on a toss, otherwise gambler B wins one point. Whoever reaches first the finishing line wins the game. Suppose that gambler A and B are m and n points away from the finishing line. Find the probability that gambler A wins the game.

Example 31 (Sum of Binomials). Suppose that $X \sim \text{Bin}(m, p)$ and $Y \sim \text{Bin}(n, p)$ are independent, then $X + Y \sim \text{Bin}(m + n, p)$.

PROOF. For any $k \in \{0, 1, \dots, m + n\}$,

$$\begin{aligned} P(X + Y = k) &= \sum_{i=0}^k P(X = i, Y = k - i) = \sum_{i=0}^k P(X = i)P(Y = k - i) \\ &= \sum_{i=0}^k C_m^i p^i (1 - p)^{m-i} \cdot C_n^{k-i} p^{k-i} (1 - p)^{n-k+i} \\ &= p^k (1 - p)^{m+n-k} \sum_{i=0}^k C_m^i C_n^{k-i} = C_{m+n}^k p^k (1 - p)^{m+n-k}. \end{aligned}$$

The last equality is due to Vandermonde identity. □

Thm 68 (Convolution). *Suppose that X, Y are independent with distribution functions F and G . Then*

$$P(X + Y \leq z) = \int_{\mathbb{R}} G(z - x) dF(x) = \int_{\mathbb{R}} F(z - y) dG(y)$$

Recall that dF, dG are notational variants for the corresponding Lebesgue-Stieltjes measure (Remark 3).

PROOF. Denote by μ, ν the distribution of X, Y . By Fubini theorem

$$\begin{aligned} P(X + Y \leq z) &= \int_{\mathbb{R}} \left(\int_{(-\infty, z-x]} d\nu(y) \right) d\mu(x) \\ &= \int_{\mathbb{R}} \left(\int_{(-\infty, z-y]} d\mu(x) \right) d\nu(y). \end{aligned}$$

The integrands respectively equal $G(z - x)$ and $F(z - y)$. □

Def 43. *The random vector (X_1, \dots, X_n) has continuous distribution if there exists a function $p \geq 0$ such that*

$$P((X_1, \dots, X_n) \in A) = \int_A p(x_1, \dots, x_n) dx_1 \cdots dx_n, \quad \forall A \in \mathcal{B}(\mathbb{R}^n).$$

The function p is called the (joint) density of (X_1, \dots, X_n) .

The definition can be equivalently³ stated as: The random vector (X_1, \dots, X_n) has continuous distribution if there exists a function $p \geq 0$ such that the joint distribution function F has

$$F(x_1, \dots, x_n) = \int_{(-\infty, x]} p(s_1, \dots, s_n) ds_1 \cdots ds_n, \quad \forall x = (x_1, \dots, x_n) \in \mathbb{R}^n,$$

where $(-\infty, x] = (-\infty, x_1] \times \cdots \times (-\infty, x_n]$.

In view of the definition, if X has density p , then its distribution function F has

$$F(x) = \int_{-\infty}^x p(s) ds, \quad \forall x.$$

³See previous section.

Example 32. X has normal distribution with parameter $\mu \in \mathbb{R}$, $\sigma > 0$, written $X \sim \mathcal{N}(\mu, \sigma^2)$, if X has density

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}.$$

Then $EX = \mu$, $\text{Var}(X) = \sigma^2$.

Thm 69. Suppose that X, Y are independent with distribution functions F and G . If X has density p_X , then $Z = X + Y$ has density

$$h(z) = \int p_X(z - y) dG(y).$$

If also Y has density p_Y , then

$$h(z) = \int p_X(z - y) p_Y(y) dy.$$

PROOF. The convolution Theorem 68 now becomes

$$P(X + Y \leq z) = \int_{\mathbb{R}} \left(\int_{-\infty}^{z-y} p_X(x) dx \right) dG(y).$$

Combining a change of variable $u = x + y$ with Fubini theorem, we obtain

$$\begin{aligned} P(X + Y \leq z) &= \int_{\mathbb{R}} \left(\int_{-\infty}^z p_X(u - y) du \right) dG(y) \\ &= \int_{-\infty}^z \left(\int_{\mathbb{R}} p_X(u - y) dG(y) \right) du \end{aligned}$$

□

Example 33. X, Y are jointly normal, denoted by $(X, Y) \sim \mathcal{N}(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$, if the joint density $p(x, y)$ is given by

$$\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \cdot \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right) + \left(\frac{y-\mu_2}{\sigma_2}\right)^2\right]\right\}.$$

Find the density of X, Y and $Z = X + Y$.

PROOF. Tedious calculations are omitted. The answers are

$$X \sim \mathcal{N}(\mu_1, \sigma_1^2), \quad Y \sim \mathcal{N}(\mu_2, \sigma_2^2).$$

and

$$Z \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + 2\rho\sigma_1\sigma_2 + \sigma_2^2).$$

□

10. 大数律 Law of large numbers

10.1. L_2 weak law.

Thm 70 (L_2 weak law). Suppose that X_1, \dots, X_n are pairwise uncorrelated with $EX_i = \mu$ and $\text{Var}(X_i) \leq C < \infty$. Let $S_n = X_1 + \dots + X_n$. Then

$$\frac{S_n}{n} \rightarrow \mu \text{ in } L_2 \text{ and probability.}$$

PROOF. We only show L_2 convergence, which will give convergence of probability via Markov inequality. Using the variance of sum formula (Theorem 67), we have

$$E \left| \frac{S_n}{n} - \mu \right|^2 = \text{Var} \left(\frac{S_n}{n} \right) = \frac{\text{Var}(S_n)}{n^2} = \frac{\sum_i \text{Var}(X_i)}{n^2} \leq \frac{C}{n} \rightarrow 0.$$

□

An important special case of the L_2 weak law is the following.

Thm 71. Suppose that X_1, \dots, X_n are i.i.d with $EX_i = \mu$ and $\text{Var}(X_i) = \sigma^2$. Let $S_n = X_1 + \dots + X_n$. Then

$$\frac{S_n}{n} \rightarrow \mu \text{ in } L_2 \text{ and probability.}$$

Below is a probabilistic proof of Weierstrass approximation theorem.

Example 34 (Bernstein polynomial). f is continuous on $[0, 1]$. Define

$$f_n(x) = \sum_{k=0}^n f\left(\frac{k}{n}\right) C_n^k x^k (1-x)^{n-k}, \quad \forall x \in [0, 1].$$

Then

$$\sup_{x \in [0, 1]} |f_n(x) - f(x)| \rightarrow 0.$$

PROOF. Observe that, if we let $S_n \sim \text{Bin}(n, x)$, then

$$f_n(x) = Ef(S_n/n).$$

Hence

$$(10.1) \quad |f_n(x) - f(x)| = |Ef(S_n/n) - f(x)| = |E[f(S_n/n) - f(x)]| \\ \leq E|f(S_n/n) - f(x)|.$$

$\forall \varepsilon > 0$, we want to bound the rightmost expectation in terms of ε . Since f is continuous and hence uniformly continuous on $[0, 1]$, we can fix δ small so that

$$|f(s) - f(t)| < \varepsilon \text{ if } |s - t| < \delta.$$

Let $M = \sup_{x \in [0,1]} f(x)$ and $A_n = \{\omega : |S_n(\omega)/n - x| < \delta\}$. Then (10.1) continues

$$\begin{aligned} |f_n(x) - f(x)| &\leq E(|f(S_n/n) - f(x)|1_{A_n}) + E(|f(S_n/n) - f(x)|1_{A_n^c}) \\ &\leq \varepsilon + 2MP(|S_n/n - x| \geq \delta) \\ &\leq_{(e_1)} \varepsilon + 2M \frac{\text{Var}(S_n/n)}{\delta^2} =_{(e_2)} \varepsilon + 2M \frac{x(1-x)}{n\delta^2} \\ &\leq \varepsilon + \frac{M}{2n\delta^2} \leq 2\varepsilon \text{ as } n \rightarrow \infty, \end{aligned}$$

where (e_1) uses Markov inequality, and (e_2) Example 29. □

LEMMA 72. *If b_n satisfies $\text{Var}(S_n)/b_n^2 \rightarrow 0$, then*

$$\frac{S_n - ES_n}{b_n} \rightarrow 0 \text{ in } L_2 \text{ and probability.}$$

PROOF. We have

$$\text{Var}\left(\frac{S_n - ES_n}{b_n}\right) = \frac{\text{Var}(S_n)}{b_n^2} \rightarrow 0.$$

□

Example 35 (Coupon collector's problem). *Suppose there are n types of coupons. You get one coupon each time you open a box of candy, and the coupon is equally likely to be any of the n types. We are interested in the time T_n to collect a complete set of coupons. Let $\tau_0^n = 0$ and*

$\tau_k^n =$ the first time we have k different coupons, $k = 1, \dots, n$.

Then

$$T_n = \tau_n^n = \sum_{k=1}^n (\tau_k^n - \tau_{k-1}^n).$$

It is readily seen that the waiting times $\{\tau_k^n - \tau_{k-1}^n\}_{k=1}^n$ between two types of coupons are independent and each has geometric distribution,

$$\tau_k^n - \tau_{k-1}^n \sim \text{Geom}\left(1 - \frac{k-1}{n}\right).$$

Example 16 tells us that

$$ET_n = \sum_{k=1}^n \left(1 - \frac{k-1}{n}\right)^{-1} = n \sum_{m=1}^n m^{-1} \approx n \log n.$$

and

$$\text{Var}(T_n) = \sum_{k=1}^n \left(1 - \frac{k-1}{n}\right)^{-2} = n^2 \sum_{m=1}^n m^{-2} \leq n^2 \sum_{m=1}^{\infty} m^{-2}.$$

Since $\sum_{m=1}^{\infty} m^{-2}$ is convergent, if we take $b_n = n \log n$, then

$$\frac{\text{Var}(T_n)}{b_n^2} \leq \frac{n^2 \sum_{m=1}^{\infty} m^{-2}}{(n \log n)^2} \rightarrow 0.$$

So Lemma 72 gives

$$\frac{T_n - n \sum_{m=1}^n m^{-1}}{n \log n} \rightarrow 0 \text{ in probability.}$$

It follows that

$$\frac{T_n}{n \log n} \rightarrow 1 \text{ in probability.}$$

This tells us that T_n is roughly $n \log n$.

Example 36 (Random permutation). A permutation of $\{1, \dots, n\}$ is a one-to-one mapping from $\{1, \dots, n\}$ to itself. There are $n!$ permutations in total. We are interested in the expected number of cycles in a randomly chosen permutation. As an example, we look at the

permutation $i \mapsto \pi(i)$,

$$\begin{array}{cccccc} i : & 1 & 2 & 3 & 4 & 5 & 6 \\ \pi(i) : & 2 & 5 & 6 & 4 & 1 & 3 \end{array}$$

Starting with 1, we follow the route of mapping

$$1 \rightarrow \pi(1) \rightarrow \pi^2(1) \rightarrow \pi^3(1) \rightarrow \cdots$$

Since $\pi^3(1) = 1$, we get a cycle

$$1 \rightarrow 2 \rightarrow 5 \rightarrow 1.$$

We use brackets to indicate cycles, so we have the first cycle (125), and the remaining cycles are (36), (4). The original permutation can now be simply written as the decomposition

$$(125)(36)(4).$$

This is indeed how random permutation algorithm works. Define the indicator random variables

$$X_{n,k} = \begin{cases} 1, & \text{a closing bracket occurs after the } k\text{-th} \\ & \text{position in the decomposition,} \\ 0, & \text{otherwise.} \end{cases}$$

Then $\{X_{n,1}, \dots, X_{n,n}\}$ are independent and

$$S_n \triangleq X_{n,1} + \dots + X_{n,n}$$

gives the total number of cycles in the permutation. We have

$$P(X_{n,k} = 1) = \frac{1}{n - k + 1}.$$

Then

$$ES_n = \sum_{k=1}^n EX_{n,k} = \sum_{k=1}^n k^{-1}$$

and noting $X_{n,k}^2 = X_{n,k}$ we have

$$\text{Var}(S_n) \leq \sum_{k=1}^n EX_{n,k}^2 = \sum_{k=1}^n EX_{n,k} = \sum_{k=1}^n k^{-1}.$$

Now applying Lemma 72 with $b_n = (\log n)^{0.5+\varepsilon}$, $\varepsilon > 0$,

$$\frac{S_n - \sum_{k=1}^n k^{-1}}{b_n} \rightarrow 0 \text{ in probability.}$$

It follows that

$$\frac{S_n}{(\log n)^{0.5+\varepsilon}} \rightarrow 1 \text{ in probability.}$$

The arbitrariness of ε indicates that $(\log n)^{0.5}$ is a threshold for the convergence.

Random permutation is commonly used in applications from coding to games, one example is the **100 prisoners riddle**. 100 prisoners, who are numbered from 1 to 100, are offered a last chance to be pardoned. At a room, there is a cupboard with 100 drawers. 100 numbers

from 1 to 100 are randomly put into these drawers. The prisoners enter the room one by one. Each prisoner can open up to 50 drawers. No communications are allowed. If every prisoner finds their numbers, all prisoners are set free, otherwise all will be sentenced. The prisoners need to figure out the best strategy to follow.

10.2. Weak law of large numbers.

Thm 73 (Weak law for triangular arrays). *Consider the triangular array of random variables $X_{n,k}$, $k = 1, \dots, n$,*

$$\begin{array}{ccccccc} & & X_{1,1} & & & & \\ & & X_{2,1} & X_{1,1} & & & \\ & & \dots & & & & \\ X_{n,1} & \dots & X_{n,k} & \dots & X_{n,n} & & \end{array}$$

Random variables in each row are pairwise independent. Let $b_n > 0$ satisfies

(10.2)

$$(i) \sum_{k=1}^n P(|X_{n,k}| > b_n) \rightarrow 0; \quad (ii) \frac{\sum_{k=1}^n \text{Var}(X_{n,k} 1_{|X_{n,k}| \leq b_n})}{b_n^2} \rightarrow 0.$$

If we set $S_n = \sum_{k=1}^n X_{n,k}$, $a_n = \sum_{k=1}^n E(X_{n,k} 1_{|X_{n,k}| \leq b_n})$, then

$$\frac{S_n - a_n}{b_n} \rightarrow 0 \text{ in probability.}$$

PROOF. Let

$$\bar{S}_n = \sum_{k=1}^n X_{n,k} 1_{|X_{n,k}| \leq b_n} \text{ and } Z_n = \frac{S_n - a_n}{b_n}.$$

We have $\forall \varepsilon > 0$,

$$\begin{aligned} P(|Z_n| > \varepsilon) &= P(|Z_n| > \varepsilon, S_n \neq \bar{S}_n) + P(|Z_n| > \varepsilon, S_n = \bar{S}_n) \\ &\leq P(S_n \neq \bar{S}_n) + P(|Z_n| > \varepsilon, S_n = \bar{S}_n). \end{aligned}$$

By assumption (i),

$$P(S_n \neq \bar{S}_n) \leq \sum_{k=1}^n P(|X_{n,k}| > b_n) \rightarrow 0.$$

Now using assumption (ii) and that $X_{n,i}, X_{n,j}, i \neq j$ are independent, we have

$$\begin{aligned} P(|Z_n| > \varepsilon, S_n = \bar{S}_n) &= P\left(\left|\frac{\bar{S}_n - a_n}{b_n}\right| > \varepsilon\right) \leq \frac{\text{Var}(\bar{S}_n)}{\varepsilon^2 b_n^2} \\ &= \frac{\sum_{k=1}^n \text{Var}(X_{n,k} 1_{|X_{n,k}| \leq b_n})}{\varepsilon^2 b_n^2} \rightarrow 0. \end{aligned}$$

The proof is complete. □

LEMMA 74. If $X \geq 0$, φ differentiable with $\varphi' > 0$ and $\varphi(0) = 0$, then

$$\int \varphi(X) dP = \int_0^\infty \varphi'(t) P(X > t) dt.$$

PROOF. An application of Theorem 56 with $Y = \varphi(X)$ gives,

$$\begin{aligned} \int Y dP &= \int_0^\infty P(Y > s) ds \\ &=_{(e1)} \int_0^\infty \varphi'(t) P(Y > \varphi(t)) dt = \int_0^\infty \varphi'(t) P(X > t) dt. \end{aligned}$$

We have performed in (e1) a change of variable $s = s(t) = \int_0^t \varphi'$. \square

Thm 75 (Weak law of large numbers). Let X_1, \dots, X_n are i.i.d with

$$(10.3) \quad xP(|X_1| > x) \rightarrow 0 \text{ as } x \rightarrow 0.$$

If we set $S_n = \sum_{k=1}^n X_k$, $\mu_n = E(X_1 1_{|X_1| \leq n})$, then

$$\frac{S_n}{n} - \mu_n \rightarrow 0 \text{ in probability.}$$

PROOF. We want to apply Theorem 73 with $X_{n,k} = X_k$ and $b_n = n$. To do this, we need to verify condition (10.2). First note that

$$\sum_{k=1}^n P(|X_{n,k}| > b_n) = nP(|X_k| > n) \rightarrow 0$$

and

$$\begin{aligned} \frac{\sum_{k=1}^n \text{Var}(X_{n,k} 1_{|X_{n,k}| \leq b_n})}{b_n^2} &= \frac{\sum_{k=1}^n \text{Var}(X_k 1_{|X_k| \leq n})}{n^2} \\ &= \frac{\text{Var}(X_1 1_{|X_1| \leq n})}{n} \leq \frac{E((X_1 1_{|X_1| \leq n})^2)}{n}. \end{aligned}$$

Thus the proof would be completed if we show that

$$\frac{E\left((X_1 1_{|X_1| \leq n})^2\right)}{n} \rightarrow 0.$$

By Lemma 74,

$$(10.4) \quad E\left((X_1 1_{|X_1| \leq n})^2\right) = \int_0^\infty 2tP(|X_1| 1_{|X_1| \leq n} > t)dt.$$

Note the integrand has

$$P(|X_1| 1_{|X_1| \leq n} > t) = P(|X_1| > t, |X_1| \leq n),$$

which gives the expression,

$$P(|X_1| 1_{|X_1| \leq n} > t) = \begin{cases} P(|X_1| > t) - P(|X_1| > n), & t < n \\ 0, & t \geq n \end{cases}$$

Hence upon substituting the above in (10.4) we obtain

$$\frac{E\left((X_1 1_{|X_1| \leq n})^2\right)}{n} \leq \frac{1}{n} \int_0^n 2tP(|X_1| > t)dt.$$

Using the assumption that $tP(|X_1| > t) \rightarrow 0$ as $t \rightarrow \infty$, we see that the RHS converges to zero, which completes the proof. \square

A sufficient condition for (10.3) is $E|X_1| < \infty$. Indeed, by dominated convergence theorem,

$$xP(|X_1| > x) \leq E(|X_1|1_{|X_1|>x}) \rightarrow 0 \text{ as } x \rightarrow \infty.$$

So $E|X_1| < \infty$ implies (10.3), but the latter is not much weaker since by Lemma 74, $\forall \varepsilon > 0$,

$$\begin{aligned} E|X|^{1-\varepsilon} &= \int_0^\infty (1-\varepsilon)t^{-\varepsilon}P(X > t)dt \\ &= \int_0^1 (1-\varepsilon)t^{-\varepsilon}P(X > t)dt + \int_1^\infty (1-\varepsilon)t^{-\varepsilon}P(X > t)dt \\ &\leq \int_0^1 t^{-\varepsilon}dt + \int_1^\infty t^{-(1+\varepsilon)}tP(X > t)dt < \infty. \end{aligned}$$

Thm 76. Let X_1, \dots, X_n are i.i.d with $E|X_1| < \infty$. If we set $S_n = \sum_{k=1}^n X_k$, $\mu = EX_1$, then

$$\frac{S_n}{n} - \mu \rightarrow 0 \text{ in probability.}$$

PROOF. Let $\mu_n = E(X_1 1_{|X_1| \leq n})$. As we have already seen that $E|X_1| < \infty$ implies (10.3), so we can employ Theorem 75 to conclude that, $\forall \varepsilon > 0$,

$$P\left(\left|\frac{S_n}{n} - \mu_n\right| > \varepsilon\right) \rightarrow 0.$$

Since $\mu_n \rightarrow \mu$ by dominated convergence theorem, we have $|\mu_n - \mu| < \varepsilon$ for large n , therefore

$$P\left(\left|\frac{S_n}{n} - \mu\right| > 2\varepsilon\right) \leq P\left(\left|\frac{S_n}{n} - \mu_n\right| > \varepsilon\right).$$

It follows that $S_n/n - \mu \rightarrow 0$ in probability. □