Notes of Probability and Stochastics

Xie Zejian Zhang Songxin

2021-02-25

Contents

0.1 Notations

$\mathbb{R}$	$(-\infty, \infty)$
$\overline{\mathbb{R}}$	$[-\infty, \infty]$
$\mathbb{R}_+$	$[0, \infty)$
$\overline{A}$	Closure of set A
A°	Interior of set A
$(x_n) \subset A$	A sequence taking value in A
2^A	The power set of A
$\mathcal{A}$	A collection of subsets in A, i.e., $\mathcal{A} \subset 2^A$
$\ker \mathcal{A}$	$\bigcap_{A \in \mathcal{A}} A$
$x_n \nearrow x$	(x_n) is increasing and converges to x.
$\sigma(\mathcal{A})$	σ-algebra generated by $\mathcal{A}$.
$\mathcal{A}_+$	Nonnegative function in $\mathcal{A}$
$\mu \ll \nu$	μ is absolutely continuous w.r.t. ν.
$\mu f = \int f d\mu = \int f(x)\mu(dx)$	integral
$f : X \to Y$	x is a function from X to Y.
$f = x \mapsto 5x$	$f(x) = 5x$
$f : X \hookrightarrow Y$	f is an embedding from X to Y.
$f(x) = O(g(x)) \iff g(x) = \Omega(f(x))$	f is bounded above by g asymptotically $\limsup_{x\to\infty} \frac{f(x)}{g(x)} < \infty$.
$f(x) = \Theta(g(x))$	f is bounded by g asymptotically both above and below .
$f(x) = o(g(x)) \iff g(x) = \omega(f(x))$	f is dominated by g asymptotically, i.e., $\lim_{x\to\infty} \frac{\lvert f(x)\rvert}{g(x)} = 0$.
$f(x) \sim g(x)$	f is equal to g asymptotically i.e. $\lim_{x\to\infty} \frac{f(x)}{g(x)} = 1$.
i_ϵ	$\mathbf{1}_{(\epsilon,\infty)}$
$d(x, A)$	$\inf\{d(x, y) : y \in A\}$
$\mathfrak{Z}$	Standard Gaussian distribution
$\mathfrak{P}_c$	Poisson distribution with mean c
s.t.	such that
w.r.t.	with respect to
r.v.	random variable

Chapter 1

Measure and integrations

1.1 Measurable space

1.1.1 σ algebra

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

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

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

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

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

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

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

1.1.2 π, λ, m systems

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

- **m-system** if closed under monotone series, that is if $(A_n) \subset \mathcal{A}$ and $A_n \nearrow A$, then $A \in \mathcal{A}$.
- $\pi-$**system** is closed under finite intersection
- $\lambda-$ **system** if

1. $\Omega \in \mathcal{A}$
2. closed under proper difference:

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

3. is an m-system.

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

Proof. For $\Rightarrow$, check:

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

For the converse:

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

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

□

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

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

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

Then we have **Monotone class theorem**:

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

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

1.1.3 Graphical illustration of different classes

$$\begin{array}{ccccc}
\sigma \text{ algebra} & \Leftrightarrow & \lambda \text{ system} & + & \pi \text{ system} \\
\Downarrow & & \Downarrow & & \\
\text{algebra} & & m \text{ system} & & \\
\Downarrow & & & & \\
\text{semialgebra} & & & & \\
\Downarrow & & & & \\
\pi \text{ system} & & & &
\end{array}$$

1.1.4 Measurable spaces

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

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

1.2 Measurable function

1.2.1 Mappings

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

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

There is some properties:

1. $f^{-1}(\Omega_2) = \Omega_1, f^{-1}(\varnothing) = \varnothing$
2. $f^{-1}(B^c) = \left[f^{-1}(B)\right]^c$
3. $$\begin{aligned} f^{-1}\left(\cup_{\gamma \in \Gamma} B_\gamma\right) &= \cup_{\gamma \in \Gamma} f^{-1}\left(B_\gamma\right) \text{ for } B_\gamma \subset \Omega_2, \gamma \in \Gamma \\ f^{-1}\left(\cap_{\gamma \in \Gamma} B_\gamma\right) &= \cap_{\gamma \in \Gamma} f^{-1}\left(B_\gamma\right) \text{ for } B_\gamma \subset \Omega_2, \gamma \in \Gamma \end{aligned}$$
   where Γ may not countable.
4. $f^{-1}(B_1 - B_2) = f^{-1}(B_1) - f^{-1}(B_2) \, \forall B_1, B_2 \subset \Omega_2$
5. $B_1 \subset B_2 \subset \Omega_2 \implies f^{-1}(B_1) \subset f^{-1}(B_2)$
6. If $\mathcal{B}$ is a $\sigma-$algebra, $f^{-1}(\mathcal{B})$ is also a $\sigma-$algebra. It's easy to check $f^{-1}(\mathcal{B})$ is closed under complement and countable union. (From properties 2 and 3)
7. If $\mathcal{C}$ is nonempty, $f^{-1}(\sigma(\mathcal{C})) = \sigma\left(f^{-1}(\mathcal{C})\right)$

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

1.2.2 Measurable functions

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

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

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

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

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

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

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

□

1.2.3 Random Variable

A r.v. X is a measurable function from $(\Omega_1, \mathcal{A})$ to $(\mathbb{R}, \mathcal{B})$. It denoted by X is $\mathcal{A}$-measurable or $X \in \mathcal{A}$

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

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

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

$$\begin{aligned} &\iff X \leq x = X^{-1}([-\infty], x) \in \mathcal{A} \quad \forall x \in \mathbb{R} \\ &\iff X \leq x = X^{-1}([-\infty], x) \in \mathcal{A} \quad \forall x \in D \end{aligned}$$

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

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

1.2.4 Construction of random variables

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

Proof. Note that

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

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

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

$\square$

Recall lemma 1.2 we have:

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

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

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

1.2.5 Limiting opts

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

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

$$\{\sup_{n\to\infty} X_n \le t\} = \bigcap_{n=1}^{\infty} \{X_n \le t\} \in \mathcal{A}$$

$$\{\inf_{n\to\infty} X_n \ge t\} = \bigcap_{n=1}^{\infty} \{X_n \ge t\} \in \mathcal{A}$$

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

$\square$

That implies

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

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

$\square$

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

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

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

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

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

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

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

Proof. Suppose

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

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

□

1.2.7 Monotone classes of function

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

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

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

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

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

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

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

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

□

Remark. If $\mathcal{M}$'s monotonicity condition only holds when f is bounded, then we can only conclude $\mathcal{A}_b \subset \mathcal{M}$ but not $\mathcal{A}_+$

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

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

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

1.3 Measure

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

It's

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

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

μ is a **measure** on $\mathcal{A}$ if

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

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

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

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

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

1.3.1 Properties of measure

1.3.1.1 Semialgebras

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

1. (**Monotonicity**): $A \subset B \implies \mu(A) \leq \mu(B)$
2. (**σ-subadditivity**):
    1. $\sum_1^\infty A_n \subset A, \quad \implies \quad \sum_1^\infty \mu(A_n) \leq \mu(A)$
    2. Moreover, if μ is a measure, then

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

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

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

1.3.1.2 Algebras

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

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

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

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

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

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

$$A = \sum C_n$$

Then

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

as $C_n \subset B_n \subset A_n$.

□

1.3.1.3 σ algebras

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

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

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

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

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

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

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

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

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

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

Proof. Take

$$\bar{\mathcal{A}} = \{A \cup N : A \in \mathcal{A}\}$$
$$\bar{\mathcal{B}} = \{A \Delta N : A \in \mathcal{A}\}$$

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

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

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

and note $\bigcup_1^{\infty} A_i \in \mathcal{A}$ and $\mu(\bigcup_1^{\infty} N_i) \leq \mu(\cup_1^{\infty} B_i) \leq \cup_1^{\infty} \mu(B_i) = 0$. Thus $\bar{\mathcal{A}}$ is closed by countable union. As for complements, note $\Omega^c = A^c \cap N^c = (A^c \cap N^c \cap B^c) \cup (A^c \cap N^c \cap B) = (A^c \cap B^c) \cup (A^c \cap N^c \cap B) \in \bar{\mathcal{A}}$.

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

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

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

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

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

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

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

□

1.3.2 Specification of measures

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

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

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

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

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

□

As consequence, we have

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

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

□

1.3.3 Atomic and diffuse measure

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

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

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

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

□

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

Proof. Let D be set of all atoms and define

$$\lambda(A) = \mu(A - D)$$
$$\nu(A) = \mu(A \cap D)$$

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

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

□

sdf

1.4 Integration

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

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

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

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

2. If $f \geq 0$,define

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

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

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

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

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

Proposition 1.1. ***(Integral over sets)***

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

*(**Absolute integrability**). $\int f$ is finite iff $\int |f|$ is finite.*
(Linearity) *If $f, g, a, b \geq 0$ or $f, g \in L^1$*

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

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

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

(Positivity) *If $f \geq 0$ a.s., then $\int f \geq 0$*
(Monotonicity) *If $f_1 \leq f \leq f_2$ a.s., then $\int f_1 \leq \int f \leq \int f_2$*
(Mean value theorem) *If $a \leq f \leq b$ a.s., then*

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

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

1.4.1 Monotone Convergence Theorem

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

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

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

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

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

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

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

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

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

Take limit both side and note g_n is increasing:

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

□

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

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

Proof. Since $f_n + g \geq 0$

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

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

□

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

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

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

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

Hence we may let g dominate h_k and get

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

□

1.4.2 Criteria for zero a.e.

Theorem 1.14 (Markov inequality)**.** *Let* $A = \{x \in \Omega : f(x) \geq M\}$*, then*

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

Proof.

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

□

Lemma 1.8. *Suppose f is measurable and non-negative and $\int f d\mu = 0$. Then $f = 0$ a.e.*

Proof. By lemma 1.14 and define $A_M = \{x \in \Omega : f(x) \geq M\}$. Consequently, $\mu(A_M) = 0$ for all $M > 0$, note $A_{\frac{1}{n}} \nearrow A_0$:

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

Hence $f = 0$ a.e.

□

Lemma 1.9. *Suppose f is integrable and $\int_A f = 0$ for all measurable A. Then $f = 0$ a.e.*

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

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

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

□

Theorem 1.15. *Suppose $f : \mathbb{R} \to \mathbb{R}$ is integrable and $\int_a^x f = 0$ for all x, then $f = 0$ a.e.*

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

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

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

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

If $G_n \searrow H$, then

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

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

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

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

□

1.4.3 Characterization of the integral

Theorem 1.16. *Let $(\Omega, \mathcal{A})$ be a measurable space and $L : \mathcal{A} \to \overline{\mathbb{R}}_+$, then there is a unique measure μ on $(\Omega, \mathcal{A})$ s.t. $L(f) = \int f$ for every $f \in \mathcal{A}_+$ iff:*

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

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

□

1.5 Transforms and Indefinite integral

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

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

Theorem 1.17. *For every $f \in \mathcal{E}$, we have $(\nu \circ h^{-1})f = \nu(f \circ h)$.*

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

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

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

□

1.5.1 Images of the Lebesgue measure

By theorem 1.17, we have a convenient tool for creating new measure from old and, we may integral a measure using Lebesgue measure:

Theorem 1.18. *Let* $(\Omega, \mathcal{A}, \mu)$ *be a standard measure space where* μ *is s-finite and* $b = \mu(\Omega)$. *Then there exists a measurable mapping* $h : ([0, b), \mathcal{B}_{[0,b]}) \to (\Omega, \mathcal{A})$ *s.t.* $\mu = \lambda \circ h^{-1}$, *where* λ *is the Lebesgue measure on* $[0, b)$.

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

1.5.2 Indefinite integrals

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

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

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

Remark. ν is a measure follows from MCT.

Proposition 1.2. *For any* $f \in \mathcal{A}_+$, $\nu f = \mu(pf)$.

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

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

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

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

□

Remark. Writing above result in an explicit notation:

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

that is:

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

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

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

1.5.3 Radon-Nikodym theorem

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

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

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

$$\int_\Omega f d\nu = \int_\Omega f p d\mu$$

1.6 Kernels and Product spaces

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

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

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

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

when fix $x \in E$, $K(x, B) = \nu(k(x, y)\mathbf{1}_B) = \mu(B)$ for some μ which is the measure on $(F, \mathcal{F})$;
when fix $B \in \mathcal{F}$, $f(x) = K(x, B)$ is measurable follows from theorem 1.4.

1.6.1 Measure-kernel-function

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

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

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

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

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

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

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

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

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

- $f(0) \Rightarrow L(f) = 0$
- If $f, g \in \mathcal{F}_+$ and $a, b \in \overline{\mathbb{R}}_+$, then:

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

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

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

  as MCT.

Hence, there exists a measure ν s.t. $L(f) = \mu(Kf) = \nu f$ as theorem 1.16. Then it suffices to show $\nu = \mu K$. Taking $f = \mathbf{1}_B$, we have $\nu(B) = \nu(\mathbf{1}_B) = \mu(K\mathbf{1}_B)$, it follows that

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

□

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

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

1.6.2 Products of kernels

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

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

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

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

1.6.3 Markov kernel

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

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

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

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

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

1.6.4 finite and bounded kernels

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

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

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

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

Remark.

$$\begin{array}{ccccc} \text{bounded} & \Longrightarrow & \sigma\text{-bounded} & \Longrightarrow & s\text{-bounded} \\ \Downarrow & & \Downarrow & & \Downarrow \\ \text{finite} & \Longrightarrow & \sigma\text{-finite} & \Longrightarrow & s\text{-finite} \end{array}$$

1.6.5 Functions on product spaces

Sections of a measurable function are measurable:

Proposition 1.3. *Let $f \in \mathcal{X} \times \mathcal{Y}$, then it's selection, $x \mapsto f(x, y)$ and $y \mapsto f(x, y)$ are measurable for each x and y.*

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

Lemma 1.10. *Let K be a s-finite kernel from $(X, \mathcal{X})$ into $(X, \mathcal{Y})$, then, $\forall f \in (\mathcal{X} \times \mathcal{Y})_+$, define*

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

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

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

Proof. By proposition 1.3, $f_x : y \mapsto f(x,y)$ is measurable in $\mathcal{F}_+$ and thus $Tf(x) = Kf_x(x)$, hence

- Linearity:

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

- Continuity from below

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

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

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

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

□

1.6.6 Measures on the product space

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

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

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

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

Proof. To see πf define a measure, check theorem 1.16, which follows from $\pi f = \mu(Tf)$ and similar properties enjoyed by T from lemma 1.10.

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

□

1.6.7 Product measures and Fubini

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

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

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

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

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

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

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

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

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

1.6.8 Finite products

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

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

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

1.6.9 Infinite products

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

Proposition 1.4. *Suppose there is $f_i : (X, \mathcal{F}) \to (\Omega_i, \mathcal{A}_i)$ for $i \in I$ and define $f(x) = (f_i(x))_{i \in I}$, then f is measurable iff each f_i is measurable.*

Chapter 2

Probability Spaces

2.1 Probability Spaces and Random Variables

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

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

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

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

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

2.1.1 Measure-theoretic and probabilistic languages

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

2.2 Distribution

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

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

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

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

when $x \in \mathbb{R}$, such $c : \mathbb{R} \to [0, 1]$ is called **distribution function(d.f.)** and is nondecreasing and right continuous. As it's increasing and bounded, the one-side limit always exists. In fact, we have

$$\begin{aligned} &c(-\infty) = \mathbb{P}\{X = -\infty\}, c(\infty) = \mathbb{P}\{X = \infty\} \\ &c(x-) = \mathbb{P}\{X < x\}, c(x) - c(x-) = \mathbb{P}\{X = x\} \end{aligned}$$

Let D be set of all atoms in μ, then D is

$$D = \{x \in \overline{\mathbb{R}} : \mathbb{P}\{X = x\} > 0\}$$

and countable by lemma 1.6. Define $D_x = D \cap [-\infty, x]$ and

$$a(x) = \sum_{y \in D_x} \mathbb{P}\{X = y\}, b(x) = c(x) - a(x)$$

for $x \in \mathbb{R}$. Then a is d.f. of measure μ_a defined by

$$\mu_a(B) = \mu(B \cap D), B \in \mathcal{B}(\overline{\mathbb{R}})$$

and b is d.f. of $\mu_b = \mu - \mu_a$. Then we decomposition μ into a purely atomic μ_a and diffuse μ_b.

2.2.1 Quantile functions

Let $q : (0, 1) \to \overline{\mathbb{R}}$ be the right continuous functional inverse of c,i.e.,

$$q(u) = \inf\{x \in \mathbb{R}, u < c(x)\}$$

Note that q is real valued iff $c(\infty) = 1$ and $c(-\infty) = 0$.

As c is right continuous, if $c(x) < u$, $q(u) > x$, but $c(x) > u$ can only implies $q(u) \leq x$.

$c(x-) \leq u$ iff $q(u) \geq x$ and $q(u-) \leq x$ iff $c(x) \geq u$.

Let λ denote the Lebesgue measure on $(0, 1)$ then $\mu = \lambda \circ q^{-1}$.

2.2.2 Joint distributions

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

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

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

thus we have

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

μ and ν are called **marginal distributions**

2.2.2.1 Independence of distribution

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

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

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

2.2.3 Stochastic process and probability laws

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

Now we can treat $X(\omega)$ as function $T \to E : t \mapsto X_t(\omega)$, thus $X : \mathcal{F} \to E^T$ is measurable as proposition 1.4 and it's a r.v. live in the same spaces as X_i and taking values in $(E^T, \mathcal{E}^T)$. It's distribution, $\mathbb{P} \circ X^{-1}$, is called **probability law** of stochastic process $\{X_t : t \in T\}$.

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

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

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

2.3 Expectation

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

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

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

2.3.1 Properties of expectation

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

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

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

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

(Linearity)

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

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

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

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

$$a\,\mathbb{P}(A) \leq \mathop{\mathbb{E}}_{A} X \leq b\,\mathbb{P}(A)$$

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

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

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

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

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

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

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

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

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

and

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

Remark. 1. If $\mathbb{P}(A) = 1$, then $\mathbb{E}\, X = \mathbb{E}_A\, X$.

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

2.3.2 Expectations and integrals

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

Theorem 2.1. *If $X \geq 0$, then*

$$\mathbb{E}\, X = \int_0^\infty \mathbb{P}\{X > x\} dx$$

Proof. Note

$$X(\omega) = \int_0^{X(\omega)} dx = \int_0^\infty \mathbf{1}_{X>x}(\omega) dx$$

then

$$\begin{aligned} \mathbb{E}\, X &= \int_\Omega X(\omega)\, \mathbb{P}(d\omega) \\ &= \int_\Omega \int_0^\infty \mathbf{1}_{X>x}(\omega) dx\, \mathbb{P}(d\omega) \\ &= \int_0^\infty \int_\Omega \mathbf{1}_{X>x}(\omega)\, \mathbb{P}(d\omega) dx \\ &= \int_0^\infty \mathbb{P}\{X > x\} dx \end{aligned}$$

□

Theorem 2.2. *Let X be a r.v. taking value in $(E, \mathcal{E})$ then*

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

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

Proof. Note $\mu = \mathbb{P} \circ X^{-1}$, then $\Longleftarrow$ follows from theorem 1.17. For $\Longrightarrow$, taking $f = \mathbf{1}_A$:

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

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

□

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

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

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

2.3.3 Laplace and Fourier transforms

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

Remark.

1. $r \mapsto \hat{\mu}_r$ is continues and decreasing on $(0,\infty)$ and note $e^{-rX} = e^{-rX}\mathbf{1}_{X<\infty} \nearrow \mathbf{1}_{X<\infty}$ as $r \searrow 0$, then $\lim_{r\to 0^+} \hat{\mu}_r = \mathbb{P}\{X < \infty\}$
2. $\hat{\mu}_r$ is also called **Moment generating function** as

$$\lim_{r\to 0^+} \frac{d^n}{dr^n}\hat{\mu}_r = (-1)^n\, \mathbb{E}\, X^n$$

   if $\mathbb{E}\, X^n < \infty$

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

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

The definition of characteristic function require taking expectation of a complex-valued r.v. Suppose Z is complex, define

$$\mathbb{E}\, Z = \mathbb{E}\, \mathfrak{R}(Z) + i\, \mathbb{E}\, \mathfrak{I}(Z)$$

Then Jensen's inequality 2.9 yields $|\mathbb{E}\, Z| \le \mathbb{E}\, |Z|$ and thus integrability of Z is equivalent to $|Z|$.

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

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

the resulting function $r \mapsto \hat{\mu}_r : \mathbb{R} \to \mathbb{C}$ is called the **Fourier transform** of μ or **characteristic function(ch.f.)** of X. We denoted it as $\varphi_X(t) = \hat{\mu}_t$.

Theorem 2.3. *Characteristic functions have following properties:*

1. *$|\varphi(t)| \le 1$ for all t and equality occurs when $t = 0$.*
2. *$\varphi(-t) = \overline{\varphi(t)}$ for all t.*
3. *φ is uniformly continues on $\mathbb{R}$.*
4. *$\varphi_{aX+b}(t) = e^{itb}\varphi_X(at)$.*
5. *A convex countable combination of ch.f.'s is a ch.f.*

Proof. 1,2,4 is clear. For 3, note

$$\begin{aligned}
|\varphi(t+h) - \varphi(t)| &= \left|\mathbb{E}(e^{i(t+h)X} - e^{itX})\right| \\
&\le \mathbb{E}\left|e^{i(t+h)X} - e^{itX}\right| \\
&= \mathbb{E}\left|e^{ihX} - 1\right| \\
&\le \mathbb{E}\,|hX|
\end{aligned}$$

Where the last inequality follows from $\left|e^{ix} - e^{iy}\right| \leq |x - y|$. If X is not integrable, DCT also guarantee the uniform continuity.

For 5, suppose their corresponding distribution is $(\mu_i)_{i\in\mathbb{N}^*}$, then the same convex combination generates a new distribution $\sum_{i\in\mathbb{N}^*}\lambda_i\mu_i$ and the corresponding ch.f.:

$$\int e^{itX} d\sum \lambda_i\mu_i = \sum \lambda_i \int e^{itX} d\mu_i = \sum \lambda_i\varphi_i(t)$$

□

Remark. As consequence, X and $-X$ have the same distribution iff φ is real-valued as $\varphi_{-X}(t) = \varphi_X(-t) = \overline{\varphi_X(t)}$.

The main reason for introducing characteristic function is if X_1 and X_2 are independent with ch.f.'s φ_1 and φ_2 then $X_1 + X_2$ has ch.f. $\varphi_1\varphi_2$ by noting

$$\mathbb{E}\, e^{it(X_1+X_2)} = \mathbb{E}(e^{itX_1}e^{itX_2}) = \mathbb{E}\, e^{itX_1}\, \mathbb{E}\, e^{itX_2} = (\varphi_1\varphi_2)(t)$$

Proposition 2.3 (The inversion formula)**.** *For interval* $(a, b) \subset \mathbb{R}$*:*

$$\lim_{T\to\infty} (2\pi)^{-1} \int_{-T}^{T} \frac{e^{-ta} - e^{-tb}}{it}\varphi(t)dt = \mu(a, b) + \frac{1}{2}\mu\{a, b\}$$

where

$$\left|\frac{e^{-ta} - e^{-tb}}{it}\right| = \left|\int_a^b e^{-itx}dx\right| \leq |b - a|$$

Also

Lemma 2.1. *For any* $x \geq 0$*:*

$$\int_0^\pi \frac{\sin t}{t}dt \geq \int_0^x \frac{\sin t}{t}dt \geq 0$$

In particular,

$$\int_0^\infty \frac{\sin t}{t}dt = \frac{\pi}{2}$$

That implies

Proposition 2.4. *Let* X, Y *taking values in* $\mathbb{R}$*, TFAE:*

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

Proof. 1 $\iff$ 2 follows from Inversion formula 2.3 immediately.

□

Remark. Similarly, we have

$$\lim_{r\to 0^+} \frac{d^n}{dr^n}\hat{\mu}_r = i^n \mathbb{E}\, X^n$$

if $\mathbb{E}\, X^n < \infty$

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

2.4 L^p-spaces and uniform integrability

Definition 2.5. Let X be a r.v. taking values in $\mathbb{R}$ with distribution μ. For p in $[1, \infty)$, define

$$\|X\|_p = (\mathbb{E}\,|X|^p)^{\frac{1}{p}}$$

and for $p = \infty$, let

$$\|X\|_\infty = \inf\{b \in R_+, |X| \le b \text{ a.s.}\}$$

Clearly $\|\cdot\|_p$ is a norm for $p \in [1, \infty]$ and

$$0 \le \|X\|_p \le \|X\|_q \le \infty$$

provided $1 \le p \le q \le \infty$ as corollary 2.1.

L^p is a vector space since:

- For any $X \in L^p$ and $a \in \mathbb{R}$,

$$\mathbb{E}\,|aX|^p = |a|^p\, \mathbb{E}\,|X|^p < \infty$$

- For any $X, Y \in L^p$, by jensen's inequality 2.9, we have

$$\left|\frac{a+b}{2}\right|^p \le (\frac{|a|+|b|}{2}^p) \le \frac{|x|^p + |b|^p}{2}$$

  that implies

$$\mathbb{E}\,|X+Y|^p \le 2^{p-1}(\mathbb{E}\,|X|^p + \mathbb{E}\,|Y|^p) < \infty$$

2.4.1 Moment inequalities

Theorem 2.4 (Young's inequality)**.** *Let f be a continues and strictly increasing function with $f(0) = 0$, then we have*

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

As consequence:

Theorem 2.5 (Holder's inequality). *Suppose that $p, q > 1$ satisfy $\frac{1}{p} + \frac{1}{q} = 1$, then*

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

Suppose $r > 1$,

$$\|XY\|_r = (\mathbb{E}\,|X^r Y^r|)^{\frac{1}{r}} \leq (\mathbb{E}\,|X^r|^p)^{\frac{1}{pr}}(\mathbb{E}\,|X^r|^q)^{\frac{1}{qr}} = \|X\|_{rp}\|Y\|_{rq}$$

That implies:

Corollary 2.1. *Suppose $p, q, r > 1$ with $\frac{1}{p} + \frac{1}{q} = \frac{1}{r}$:*

$$\|XY\|_r \leq \|X\|_p\|Y\|_q$$

Theorem 2.6 (Cauchy-Schwarz inequality).

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

And:

Theorem 2.7 (Lyapunov's inequality).

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

Theorem 2.8 (Minkowski's inequality). $\forall p \in [1, \infty]$,

$$\|\sum X\|_p \leq \sum \|X\|_p$$

Theorem 2.9 (Jensen's inequality). *Let ψ be convex, that is, $\forall \lambda \in (0,1), x, y \in \mathbb{R}$:*

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

Then

$$\psi(\mathbb{E}\,X) \leq \mathbb{E}[\psi(X)]$$

Theorem 2.10 (Chebyshev inequality). *If g is strictly increasing and positive on $\mathbb{R}_+$, $g(x) = g(-x)$, and X is a r.v. s.t. $\mathbb{E}(g(X)) < \infty$, then $\forall a > 0$*

$$\mathbb{P}(|X| \geq a) \leq \frac{\mathbb{E}\,g(X)}{g(a)}$$

2.4.2 Uniform integrability

Lemma 2.2. *Let X taking values in $\mathbb{R}$, then it's integrable iff*

$$\lim_{b\to\infty} \mathbb{E}\,|X|\mathbf{1}_{|X|>b} = 0$$

Proof. $\implies$ is follows from theorem 1.12 as $|X|\mathbf{1}_{|X|>b} \searrow 0$. Conversely, taking $b = c \gg 1$ s.t. $\mathbb{E}\,|X|\mathbf{1}_{|X|>c} \le 1$ and then

$$\mathbb{E}\,|X| \le \mathbb{E}(c + |X|\mathbf{1}_{|X|>c}) \le c + 1 < \infty$$

□

Definition 2.6. A collection of r.v. taking values in $\mathbb{R}$, $\mathcal{K}$, is said to **uniformly integrable** if

$$k(b) = \sup_{X \in \mathcal{K}} \mathbb{E}\,|X|\mathbf{1}_{|X|>b} \to 0$$

as $b \to \infty$.

Remark.
1. If $\mathcal{K}$ is finite and each of $\mathcal{K}$ is integrable then $\mathcal{K}$ is uniformly integrable.
2. If $\mathcal{K}$ is dominated by an integrable Y then it's uniformly integrable.
3. Uniform integrability implies L^1-boundedness: $\mathcal{K} \subset L^1$ and $\sup_{\mathcal{K}} \mathbb{E}\,|X| < \infty$. That follows from

$$\begin{aligned}\mathbb{E}\,|X| &\le \mathbb{E}\left(b + \mathbb{E}\,X\mathbf{1}_{|X|>b}\right)\\ &= b + \mathbb{E}\,X\mathbf{1}_{|X|>b}\\ &\le b + k(b)\end{aligned}$$

holds for each $X \in \mathcal{K}$.

L^1 boundedness is not sufficient for uniform integrability. In fact, we need:

Theorem 2.11. *A collection of r.v. taking values in $\mathbb{R}$, $\mathcal{K}$, is uniformly integrable iff it's L^1-bounded and $\forall \epsilon > 0, \exists \delta > 0$ s.t. $\forall F \in \mathcal{F}$:*

$$\mathbb{P}(F) \le \delta \implies \sup_{X \in \mathcal{K}} \mathbb{E}\,|X|\mathbf{1}_F \le \epsilon$$

Proof. We may assume $X \ge 0$ by obvious reason. Note $X\mathbf{1}_F \le b\mathbf{1}_F + X\mathbf{1}_{X>b}$ for each F and b, take expectation:

$$\sup_{X \in \mathcal{K}} \mathbb{E}\,X\mathbf{1}_F \le b\,\mathbb{P}(F) + k(b)$$

then $\implies$ is immediately as $k(b)$ can be arbitrary small.

Conversely, by Markov's inequality 1.14:

$$\sup_{X \in \mathcal{K}} \mathbb{P}\{X > b\} \le \frac{1}{b} \sup_{X \in \mathcal{K}} \mathbb{E}\,X = \frac{k(0)}{b}$$

that suggests we may choose b s.t. $\mathbb{P}\{X > b\}$ arbitrary small, and thus $\sup \mathbb{E}_F X$ arbitrary small, taking $H = \{X > b\}$, then we have definition of uniformly integrability exactly.

□

However, L^p boundedness when $p > 1$ implies uniform integrability.

Lemma 2.3. *Suppose there is a borel $f : \mathbb{R}_+ : \overline{\mathbb{R}}_+$ s.t. $f(x) = \omega(x)$ and*

$$\sup_{X \in \mathcal{K}} \mathbb{E} f \circ |X| < \infty$$

then $\mathcal{K}$ is uniformly integrable.

Proof. Again we may assume $X \geq 0$ and it's sufficient to assume $f \geq 1$, let $g(x) = \frac{x}{f(x)}$ and note

$$X\mathbf{1}_{X>b} = f \circ Xg \circ X\mathbf{1}_{X>b} \leq f \circ X \sup_{x>b} g(x)$$

let $c = \sup_{X \in \mathcal{K}} f \circ X \leq \infty$, we have

$$k(b) \leq c \sup_{x>b} g(x)$$

it follows $\lim_{b \to \infty} k(b) = 0$ as $\lim_{x \to \infty} g(x) = 0$

□

And the converse is also true:

Theorem 2.12. *Using notations above, TFAE:*

1. *$\mathcal{K}$ is uniformly integrable.*
2. *$h(b) = \sup_{\mathcal{K}} \int_b^\infty \mathbb{P}\{|X| > y\} \to 0$ as $b \to \infty$.*
3. *$\sup_{\mathcal{K}} \mathbb{E} f \circ |X| < \infty$ for some increasing convex f on $\mathbb{R}_+$ s.t. $f(X) = \omega(x)$.*

Proof. Assume X is positive and it suffices to show $1 \implies 2 \implies 3$.

$1 \implies 2$. $\forall X \in \mathcal{K}$,

$$\begin{aligned} \mathbb{E} X\mathbf{1}_{X>b} &= \int_0^\infty \mathbb{P}\{X\mathbf{1}_{X>b} > y\} dy \\ &= \int_0^\infty \mathbb{P}\{X > b \vee y\} dy \\ &\geq \int_b^\infty \mathbb{P}\{X > y\} dy \end{aligned}$$

thus $k(b) \geq h(b)$ and claim follows.

$2 \implies 3$ follows from construction and omitted.

□

2.5 Information and determinability

2.5.1 σ algebra generated by r.v.

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

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

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

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

1. $$\sigma(X_i) = \sigma\left(X_i^{-1}(\mathcal{B})\right) = X_i^{-1}(\mathcal{B}) = \{X_i \in \mathcal{B}\}$$ $$\sigma(X_1, \dots, X_n) = \sigma\left(\cup_{i=1}^n X_i^{-1}(\mathcal{B})\right) = \sigma\left(\cup_{i=1}^n \sigma(X_i)\right)$$

2. $$\sigma(X_1) \subset \sigma(X_1, X_2) \subset \dots\dots \subset \sigma(X_1, \dots, X_n)$$ $$\sigma(X_1, X_2, \dots..) \supset \sigma(X_2, X_3, \dots..) \supset \dots\dots \supset \sigma\left(X_n, X_{n+1}, \dots..\right)$$

3. $\bigcap_1^\infty \sigma(X_n, X_{n+1}, \cdots)$ is the tail σ algebra of $X_{1:}$

In view of 1.4:

Proposition 2.5. *If $X = (X_t)_{t\in T}$, then $\sigma X = \sigma\{X_t : t \in T\}$*

Theorem 2.13 (Doob-Dynkin lemma)**.** *Let X be a r.v. taking values in space $(E, \mathcal{E})$. A mapping $V : \Omega \to \overline{\mathbb{R}}$ belongs to σX iff $V = f(X)$ for some $f \in \mathcal{E}$.*

Proof. $\Longleftarrow$ is immediately as measurable functions of measurable functions are measurable.

$\Longrightarrow$. Let $\mathcal{M} = \{V : V = f(X)\}$, then it's a monotone class and claim follows from theorem 1.5.

$\square$

Putting $X = (X_1, X_2, \dots)$ lead to

Corollary 2.2. *Suppose $(X_n)_{n\in\mathbb{N}^*}$ are all r.v., then $V : \Omega \to \overline{\mathbb{R}}$ belongs to $\sigma\{X_n : n \in \mathbb{N}^*\}$ iff $V = f(X_1, X_2, \dots)$ for some $f \in \prod_{i\in\mathbb{N}^*} \mathcal{E}_i$.*

This can be generalized to uncountable case:

Proposition 2.6. *Suppose $(X_t)_{t\in T}$ is family of r.v. then $V : \Omega \to \overline{\mathbb{R}}$ belongs to $\sigma\{X_t : t \in T\}$ iff there exist countable $(t_n) \subset T$ and a function $f \in \prod_{(t_n)} \mathcal{E}_{t_n}$ s.t. $V = f(X_{t_1}, X_{t_2}, \dots)$.*

Definition 2.7. Suppose X and Y are r.v., then we say X **determines** Y if $Y = f \circ X$ for some measurable f. σX is called **information** as it contains all determined variables w.r.t. X.

2.5.2 Filtrations

Definition 2.8. A filtration is a filter with a total inclusion order where elements are all σ-algebra and denoted as $\mathcal{F} = \{\mathcal{F}_t : t \in T\}$ where $\mathcal{F}_t \subset \mathcal{F}_t$ provided $s < t$.

Our aim is to approximate eternal variables by known r.v.:

Theorem 2.14. *Let $\mathcal{F} = \{\mathcal{F}_n : n \in \mathbb{N}\}$ be a filtration and $\mathcal{F}_\infty = \sigma(\bigcup_{t\in T} \mathcal{F}_t) = \bigcup_{t\in T} \mathcal{F}_t$. For bounded $V \in \mathcal{F}_\infty$ there are sequence of bounded $V_n \in \mathcal{F}_n, n \in \mathbb{N}$, s.t.:*

$$\lim_{n\to\infty} \mathbb{E}\,|V_n - V| = \lim_{n\to\infty} \mathbb{E}\,V_n - \mathbb{E}\,V = 0$$

Proof. Let $\mathcal{M}_b \subset \mathcal{F}_\infty$ be collection of bounded variables can be approximated. It follows that $\mathcal{M}_b$ is a monotone class and claim follows from theorem 1.5.

□

2.6 Independence

Suppose $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and let $(\mathcal{F}_i)_{i\in I}$ be a finite family sub-σ-algebra of $\mathcal{A}$, then $\{\mathcal{F}_i : i \in I\}$ is called **independency** if

$$\mathbb{E}\prod_{i\in I} V_n = \prod_{i\in I} \mathbb{E}\,V_i$$

for all positive $V_i \in \mathcal{F}_i$ respectively.

If I is arbitrary, then $\{\mathcal{F}_t : t \in I\}$ is independency if every finite subset of it is so.

2.6.1 Independence of σ-algebras

Lemma 2.4. *Suppose $(\mathcal{F})_{i\in S}$ be a finite family of sub-σ-algebras, let $\mathcal{C}_i$ be a π-system that generates $\mathcal{F}_i$ respectively, then $\{\mathcal{F}_i : i \in I\}$ are independent iff:*

$$\mathbb{P}(\bigcap_{i\in I} A_i) = \prod_{i\in I} \mathbb{P}(A_i)$$

for any $A_i \in \mathcal{C}_i \cup \{\Omega\}$ respectively.

Proof. $\implies$ is immediately by taking $V_i = \mathbf{1}_{A_i}$. For $\impliedby$, clearly the equality holds for all $A_i \in \mathcal{F}_i$ respectively in view of theorem 1.8. It follows that indicator $r.v.$ are independent and we can extend to general V_i by theorem 1.4 and theorem 1.10.

□

2.6.2 Independence of collection

Proposition 2.7. *Every partition of independency is an independency: let $\{\mathcal{F}_t : t \in T\}$ be an independency and $(T_i)_1^\infty$ be a partition of T then $\{\mathcal{F}_{T_i}\}_i^\infty$ is an independency.*

Proof. Let $\mathcal{C}_i$ be all events having the form $\bigcap_S A_s$ where $A_s \in \bigcup_{t\in T_i} \mathcal{F}_t$, then they are π-systems contains Ω and generates $\mathcal{F}_{T_i}$. Then $\{\mathcal{F}_{T_i} : 1 \leq i\}$ is an independency follows from lemma 2.4 and $\{\mathcal{F}_t : t \in T\}$ is an independency. □

A collection of objects are said to be pairwise independent if every pair of them is an independency. Though it's weaker than mutually independent, we can check independency by respected checking pairwise independency.

Lemma 2.5. *Countable collection of sub-σ-algebras $\{\mathcal{F}_i\}_1^\infty$ are independent iff $\mathcal{F}_{\{1\leq i\leq n\}}$ and $\mathcal{F}_{n+1}$ are independent for all $n \geq 1$.*

Proof. $\implies$ is immediate from 2.7. For $\impliedby$, let $\mathcal{G}_n = \sigma(\bigcup_i^n \mathcal{F}_i)$ and $A_i \in \mathcal{F}_i$ respectively for $1 \leq i \leq m$ note:

$$\bigcap_1^{m-1} A_i \in \mathcal{G}_{m-1}$$

thus we can repeat apply lemma 2.4 and finally get what we need for apply lemma 2.4. □

2.6.3 Independence of r.v.'s

Lemma 2.6. *The r.v.'s $X_1, \ldots, X_n$ are independent iff*

$$\mathbb{E} \prod_{i=1}^{\infty} f_i \circ X_i = \prod_{i=1}^{\infty} \mathbb{E}\, f_i \circ X_i$$

for all $f_i \in \mathcal{E}_i$ respectively.

Proof. Clearly from $f \circ X \in \sigma X$ □

Let π be joint distribution of $X_1, \ldots, X_n$ and let $\mu_1, \ldots, \mu_n$ be corresponding marginals. Then the equality becomes

$$\int_{\prod_{i=1}^n E_i} \prod_{i=1}^n f_i(x_i) d\pi = \prod_{i=1}^n \int_{E_i} f_i(x_i) d\mu_i$$

and that suggests $\pi = \prod_{i=1}^n \mu_i$.

Proposition 2.8. *The random variables $X_1, \ldots, X_n$ are independent iff their joint distribution is the product of their marginal distributions.*

In view of determined variables are in σX, we have

Proposition 2.9. *Measurable functions of independent r.v.'s are independent.*

2.6.4 Sum of independent r.v.'s

Let real valued r.v.'s X and Y with distribution μ and ν are independent. The distribution of $X+Y$ denoted as $\mu * \nu$ and given by

$$(\mu * \nu)f = \mathbb{E}\, f(X+Y) = \iint f(x+y) d\nu d\mu$$

This distribution $\mu * \nu$ is called **convolution** and can be extend to any number of distributions easily.

2.6.5 Kolmogorov's 0-1 law

Definition 2.9. Let $(\mathcal{G}_n)$ be a sequence of sub-σ-algebras. We may treat $\mathcal{G}_n$ as the information revealed by the nth trial of an experiment. Then $\mathcal{J}_n = \sigma(\bigcup_{m>n} \mathcal{G}_m)$ is information after n and $\mathcal{J} = \bigcap_n \mathcal{J}_n$ is that about **remote future** and called **tail-σ-algebra**.

The sets of which are called **tail events**, and functions on which are **tail functions**.

Theorem 2.15 (Kolmogorov's 0-1 law)**.** *Tail events of independent* $(\mathcal{G}_i)_1^\infty$ *have probability* 0 *or* 1.

Proof. By proposition 2.7, $\{\mathcal{G}_i\}_1^n \cup \{\mathcal{J}_n\}$ is independency for each n which implies so is $\{\mathcal{G}_i\}_1^n \cup \{\mathcal{J}\}$ as $\mathcal{J} \subset \mathcal{J}_n$ and thus so is $\{\mathcal{G}_i\}_1^\infty \cup \{\mathcal{J}\}$ by definition, that implies $\{\mathcal{J}, \mathcal{J}_0\}$ is an independency by proposition 2.7 again and so is $\{\mathcal{J}, \mathcal{J}\}$ by noting $\mathcal{J} \subset \mathcal{J}_0$. Finally, for any event $A \in \mathcal{J}$, we have:

$$\mathbb{P}(A \cap A) = \mathbb{P}(A)^2 \implies \mathbb{P}(A) = 0 \text{ or } 1$$

as lemma 2.4.

□

Corollary 2.3. *Tail function of independent r.v.'s are degenerate a.s.*

Proof. Note that $Y \leq c$ is tail events.

□

By above corollary, we can see that $\limsup_n X_n$ and $\liminf_n X_n$ are degenerate a.s.

2.6.6 Hewitt-Savage 0-1 law

Definition 2.10. A **finite permutation** of $\mathbb{N}$ is a map $\pi : \mathbb{N} \to \mathbb{N}$ s.t. $\pi(n) = n$ for all but finite exception. For such permutation π, we write

$$X \circ \pi = \{X_{\pi(i)} : i \in \mathbb{N}\}$$

for countable $X = (X_1, X_2, \dots)$. Variable is said to be permutation invariant if $V \circ \pi = V$ for any π and event is said to be so if its indicator is such.

The collection of all permutation invariant events is a σ-algebra which contains the tail-σ-algebra of X.

The following theorem generalized kolmogorov 0-1 law 2.15 in i.i.d. cases.

Theorem 2.16 (Hewitt-Savage 0-1 law)**.** *Suppose* $(X_i)_{i\in\mathbb{N}}$ *are i.i.d., then every permutation invariant event has probability* 0 *or* 1 *and every permutation invariant r.v. is degenerate a.s..*

Proof. It's sufficient to show that if $V : \Omega \to [0,1]$ is permutation invariant in $\mathcal{F}_\infty$, then $\operatorname{Var}[V] = \mathbb{E}\, V^2 - (\mathbb{E}\, V)^2 = 0$. For such V, there exist $\{V_n : n \in \mathbb{N}\}$ and also bounded in $[0,1]$ by theorem 2.14 s.t.:

$$\lim_{n\to\infty} \mathbb{E}\,|V - V_n| = \lim_{n\to\infty} \mathbb{E}\, V_n - \mathbb{E}\, V = 0$$

As $(X_i)_{i\in\mathbb{N}}$ are i.i.d., V and $V \circ \pi$ share the same distribution and thus same expectation:

$$\begin{aligned}
\mathbb{E}\,|V - V_n| &= \mathbb{E}\,|(V - V_n) \circ \pi| \\
&= \mathbb{E}\,|V \circ \pi - V_n \circ \pi| \\
&= \mathbb{E}\,|V - V_n \circ \pi|
\end{aligned}$$

Note we can taking π s.t. V and $V_n \circ \pi$ are independent when n is fixed, then

$$\mathbb{E}\, V_n \cdot V_n \circ \pi = (\mathbb{E}\, V_n)^2$$

which in turn show that

$$\begin{aligned}
|\,\mathbb{E}\, V^2 - (\mathbb{E}\, V_n)^2| &= |\,\mathbb{E}(V^2 - V_n \cdot V_n \circ \pi)| \\
&\leq \mathbb{E}\,|V^2 - V_n \cdot V_n \circ \pi| \\
&\leq 2\,\mathbb{E}\,|V - V_n| \to 0
\end{aligned}$$

where the final step followed by noting:

$$|V^2 - V_n \cdot V_n \circ \pi| = |(V - V_n)V + (V - V_n \circ \pi)V_n| \leq |V - V_n| + |V - V_n \circ \pi|$$

$\square$

Chapter 3

Convergence

3.1 Convergence of Real Sequences

Suppose $(x_i)_{i\in\mathbb{N}^*} \subset \mathbb{R}$, then $(x_i) \to x$ iff $|x_i - x| \to 0$ and the classical statement for convergence is the same as

$$\sum_{i=1}^{\infty} i_\epsilon(|x_i - x|) < \infty \iff \limsup_{n\to\infty} i_\epsilon(|x_i - x|) = 0 \iff i_\epsilon(|x_i - x|) \to 0$$

holds for all $\epsilon > 0$.

3.1.1 Cauchy criterion

Following are useful for determining convergence.

Proposition 3.1 (Cauchy criterion)**.** *Sequence (x_n) converges iff*

$$\lim_{m,n\to\infty} |x_m - x_n| = 0$$

Proposition 3.2. *If there exists a positive sequence (ϵ_n) s.t.*

$$\sum_n \epsilon_n < \infty, \sum_n i_{\epsilon_n}(|x_{n+1} - x_n|) < \infty$$

then (x_n) is convergent.

3.1.2 Subsequence

Definition 3.1. Let $(x_i)_{i\in\mathbb{N}^*}$ be a sequence. Then $(x_{k_i})_{i\in\mathbb{N}^*}$ is a subsequence of $(x_i)_{i\in\mathbb{N}^*}$ if $(k_i)_{i\in\mathbb{N}^*} \subset \mathbb{N}$ if it's increasing with $\lim_{i\to\infty} k_i = \infty$. Regarding $\mathbb{N}$ as a sequence then $(k_i)_{i\in\mathbb{N}^*}$ is a subsequence of $\mathbb{N}$. Denoted $(k_i)_{i\in\mathbb{N}^*}$ as N, we can write $(x_i)_{i\in N}$ for $(x_{k_i})_{i\in\mathbb{N}^*}$ and we say $(x_i)_{i\in\mathbb{N}^*}$ converges along N to x if $\lim_{i\to\infty} x_{k_i} = x$.

Proposition 3.3. *Subsequence converges to* $\limsup_{n\to\infty} x_n$ *as a maximum and* $\liminf_{n\to\infty} x_n$ *as a minimum, the sequence converges iff they meet, i.e., every subsequence converges to the same point.*

Proposition 3.4. *If every subsequence of* (x_n) *has a further subsequence converges to* x*, then* $x_n \to x$*.*

Following is useful in proving LLN.

Lemma 3.1. *Let* $(x_i)_{i\in\mathbb{N}^*} \subset \overline{\mathbb{R}}_+$ *and put* $\overline{x}_n = \frac{\sum_{i=1}^n x_i}{n}$*. Let* $N = (n_k)$ *is a subsequence of* $\mathbb{N}$ *with* $\lim_{k\to\infty} \frac{n_{k+1}}{n_k} = r \geq 1$*. If the sequence* (x_i) *converges along* N *to* x*, then*

$$\frac{x}{r} \leq \liminf_{n\to\infty} \overline{x}_n \leq \limsup_{n\to\infty} \overline{x}_n \leq rx$$

Proof. For $n_k \leq n < n_{k+1}$, note that

$$\frac{n_k}{n_{k+1}}\overline{x}_{n_k} = \frac{\sum_{n_k} x_i}{n_{k+1}} \leq \overline{x}_n \leq \frac{\sum_{n_{k+1}} x_i}{n_k} = \frac{n_{k+1}}{n_k}\overline{x}_{n_{k+1}}$$

Take limit each side we have result desired.

□

3.1.3 Diagonal method

Lemma 3.2. *Suppose there is a countable family of bounded sequence* $\{S_i : i \in \mathbb{N}\}$*, then there exists a subsequence* N *of* $\mathbb{N}$ *s.t. each of them converges along* N*.*

Proof. As each S_i is bounded, we can pick N_1 s.t. S_1 converges along N_1, then consider S_2 along N_1 as a new sequence there exists $N_2 \subset N_1$ s.t. which converges along N_2. Thus for each $m < \infty$, we have S_i converges along N_m for $i \leq m$.

Now let n_m be the mth entry of N_m and define $N = (n_1, n_2, \dots)$, one can check it's tail is a subsequence of each N_i and thus S_i converges along N for each i.

□

Remark. If the sequence $(N_i)_{i\in\mathbb{N}^*}$ is written as

$$\begin{bmatrix} N_1 \\ N_2 \\ \vdots \end{bmatrix}$$

then N is precisely the diagonal of above matrix, and that's why it called diagonal method.

Following is an application of some importance.

3.1.4 Helly's Theorem

Theorem 3.1 (Helly's theorem)**.** *Suppose $(c_i)_{i\in\mathbb{N}^*}$ is a sequence of d.f.'s, then there exists a subsequence $N \subset \mathbb{N}^*$ $(c_i)_{i\in N}$ and a d.f. c s.t. $\lim_{i\in N} c_i(t) = c(t)$ at which t is continuous.*

Proof. Treat $(c_i(r))_{i\in\mathbb{N}^*}$ as a sequence and r is taken as an enumeration of $\mathbb{Q}$. Consider the subsequence $N \subset \mathbb{N}$ from lemma 3.2, we claim that $(c_i)_{i\in N}$ is exactly $(b_i)_{i\in\mathbb{N}^*}$, recall that $b(r) = \lim_{n\to\infty} b_n(r)$ exists for each $r \in \mathbb{Q}$.

For each $t \in \mathbb{R}$, define

$$c(t) = \inf\{b(r) : r \in \mathbb{Q} \text{ and } r > t\}$$

One can check c is a d.f. Then suppose c is continuous at t, for any $\epsilon > 0$ there is a $s < t$ s.t. $c(s) > c(t) - \epsilon$ and there is a rational $r > t$ s.t. $b(r) < c(t) + \epsilon$ by definition. Pick rational q s.t. $s < q < t < r$, we have

$$c(t) - \epsilon < c(s) \leq b(q) \leq b(r) < c(t) + \epsilon$$

note:

$$\liminf_{n\to\infty} b_n(t) \geq \liminf_{n\to\infty} b_n(q) = b(q)$$
$$\limsup_{n\to\infty} b_n(t) \leq \limsup_{n\to\infty} b_n(r) = b(r)$$

thus they are sandwiched by $c(t) - \epsilon$ and $c(T) + \epsilon$ and thus agree at $c(t)$ and it follows that $\lim_{n\to\infty} b_n(t) = c(t)$.

□

3.1.5 Kronecker's Lemma

Following relates convergence of averages and convergence.

Lemma 3.3. *Suppose $(x_i)_{i\in\mathbb{N}^*} \subset \mathbb{R}$ and $(a_i)_{i\in\mathbb{N}^*} \nearrow \infty$ be strictly positive. Put $y_n = \sum_{i=1}^n \frac{x_i}{a_i}$. If $(y_i)_{i\in\mathbb{N}^*}$ converges, then*

$$\lim_{n\to\infty} \frac{\sum_{i=1}^n x_i}{a_n} = 0$$

Proof. Put $a_0 = y_0 = 0$. Note $x_i = (y_i - y_{i-1})a_i$ and

$$\sum_{i=1}^n x_i = \sum_{i=0}^{n-1} (a_{i+1} - a_i)(y_n - y_i)$$

By Cauchy criterion 3.1, there exists k for any ϵ s.t. $|y_n - y_m| \leq \epsilon$ for all

$n, m \geq k$ and thus

$$\begin{aligned}\left|\sum_{i=1}^{n} x_i\right| &= \left|\sum_{i=0}^{n-1}(a_{i+1}-a_i)(y_n-y_i)\right| \\ &\leq \left|\sum_{i=k}^{n-1}(a_{i+1}-a_i)|y_n-y_i|\right| + \left|\sum_{i=0}^{k-1}(a_{i+1}-a_i)|y_n-y_i|\right| \\ &\leq (a_n-a_k)\epsilon + \left|\sum_{i=0}^{k-1}(a_{i+1}-a_i)|y_n-y_i|\right| \\ &\leq a_n\epsilon + \left|\sum_{i=0}^{k-1}(a_{i+1}-a_i)|y_n-y_i|\right|\end{aligned}$$

where the second term is finite and thus dominated by a_n, that implies $\lim_{n\to\infty} \frac{|\sum_{i=1}^{n} x_i|}{a_n} \to 0$ and then claim follows.

□

3.2 Almost Sure Convergence

Throughout, $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space and $(X_i)_{i\in\mathbb{N}^*}$ is a sequence of real-valued r.v.'s.

Sequence of r.v.'s (X_i) is said to be converges if $(X_i(\omega))$ is so for all $\omega \in \Omega$.

As $\liminf_{n\to\infty} X_n$ and $\limsup_{n\to\infty} X_n$ are r.v.'s, the set

$$\Omega_0 = \{\omega : \liminf_{n\to\infty} X_n = \limsup_{n\to\infty} X_n\}$$

is an event. Then $(X_i)_{i\in\mathbb{N}^*}$ is converges a.s. iff $\mathbb{P}(\Omega_0) = 1$.

3.2.1 Borel-Cantelli lemmas

Definition 3.2. Suppose $(A_i)_{i\in\mathbb{N}^*}$ is sequence of events, then **infinite often(i.o.)** and **ultimately(ult.)** are defined by:

$$\{A_n,\ i.o.\} = \limsup_n A_n = \bigcap_{n=1}^{\infty}\bigcup_{m=n}^{\infty} A_m = \lim_{n\to\infty} \sup\{A_{n:\infty}\} = \lim_{n\to\infty}\bigcup_{m=n}^{\infty} A_m$$

$$\{A_n,\ ult.\} = \liminf_n A_n = \bigcup_{n=1}^{\infty}\bigcap_{m=n}^{\infty} A_m = \lim_{n\to\infty} \inf\{A_{n:\infty}\} = \lim_{n\to\infty}\bigcap_{m=n}^{\infty} A_m$$

Remark. By intuition, we have:

1. $\{A_n \text{ ult.}\}^c = \{A_n^c \text{ i.o.}\}$
2. $\mathbf{1}_{\{A_n \text{ i.o.}\}} = \{\mathbf{1}_{A_n} \text{ i.o.}\}$, $\mathbf{1}_{\{A_n \text{ ult.}\}} = \{\mathbf{1}_{A_n} \text{ ult.}\}$
3. $\{A_n \text{ i.o.}\} = \{\omega : \sum_{n=1}^{\infty} \mathbf{1}_{A_n} = \infty\}$, $\{A_n \text{ ult.}\} = \{\omega : \sum_{n=1}^{\infty} \mathbf{1}_{A_n^c} < \infty\}$
4. $\{A_n \text{ i.o.}\} \subset \{A_n \text{ ult.}\}$

5. $\mathbb{P}(\liminf_{n\to\infty} A_n) \le \liminf_{n\to\infty} \mathbb{P}(A_n) \le \limsup_{n\to\infty}(A_n) \le \mathbb{P}(\limsup_{n\to\infty} A_n)$

Borel-Cantelli lemma implies some sufficient conditions for a.s. convergence.

Lemma 3.4 (Borel-Cantelli lemma)**.** *Let $(A_i)_{i\in\mathbb{N}^*}$ be a sequence of events. Then*

$$\sum_{i=1}^{\infty} \mathbb{P}(A_i) < \infty \implies \sum_{i=1}^{\infty} \mathbf{1}_{A_i} < \infty \text{ a.s.} \iff \mathbb{P}\{A_n \text{ i.o.}\} = 0$$

If (A_i) are independent, then

$$\sum_{i=1}^{\infty} \mathbb{P}(A_i) = \infty \implies \sum_{1=1}^{\infty} \mathbf{1}_{A_i} = \infty \text{ a.s.} \iff \mathbb{P}\{A_n \text{ i.o.}\} = 1$$

Proof. By the MCT 1.10,

$$\mathbb{E}\sum_{i=1}^{\infty} \mathbf{1}_{A_i} = \sum_{i=1}^{\infty} \mathbb{E}\, 1_{A_i} = \sum_{i=1}^{\infty} \mathbb{P}(A_i) < \infty$$

and the claim follows as remark 2 in **Properties of expectation**.

If (A_i) are independent, noting $1 - x \le e^{-x}$ for all $x \in \mathbb{R}$, it's sufficient to show that $0 \le 1 - \mathbb{P}\{A_n \text{ i.o.}\} \le 0$:

$$\begin{aligned}
0 &\le 1 - \mathbb{P}\{A_n \text{ i.o.}\} = \mathbb{P}\{A_n^c \text{ ult.}\} \\
&= \mathbb{P}(\liminf_{n\to\infty} A_n^c) = \lim_{n\to\infty} \mathbb{P}(\inf_{m\ge n} A_m^c) \\
&= \lim_{n\to\infty} \mathbb{P}(\bigcap_{m=n}^{\infty} A_m^c) = \lim_{n\to\infty}\lim_{r\to\infty} \mathbb{P}(\bigcap_{m=n}^{r} A_m^c) \\
&= \lim_{n\to\infty}\lim_{r\to\infty} \prod_{m=n}^{r} [1 - \mathbb{P}(A_m)] \text{ (by independence)} \\
&\le \lim_{n\to\infty}\lim_{r\to\infty} \prod_{m=n}^{r} [e^{-\mathbb{P}(A_i)}] = \lim_{n\to\infty}\lim_{r\to\infty} [e^{-\sum_{m=n}^{r} \mathbb{P}(A_m)}] \\
&= \lim_{n\to\infty} e^{-\sum_{m=n}^{\infty} \mathbb{P}(A_m)} = \lim_{n\to\infty} 0 = 0
\end{aligned}$$

□

3.2.2 Convergence a.s. theorem

Theorem 3.2. *Suppose $(X_i)_{i\in\mathbb{N}^*}$ is sequence of r.v.'s, TFAE:*

1. *$X_n \to X$ a.s.*
2. *For every $\epsilon > 0$,*

$$\sum_{i=1}^{\infty} i_\epsilon \circ |X_n - X| < \infty \text{ a.s.} \iff i_\epsilon \circ |X_n - X| \to 0 \text{ a.s.}$$

3. *For every $\epsilon > 0$,*
$$\mathbb{P}\{|X_n - X| \geq \epsilon \ i.o.\} = 0$$

Proof. 1 $\iff$ 2 follows from the equivalent statement of convergence stated at the beginning.

2 $\iff$ 3 follows from remark 1 and 3 in definition 3.2.

□

3.2.3 Cauchy criterion for convergence a.s.

Theorem 3.3. *Suppose $(X_i)_{i\in\mathbb{N}^*}$ is sequence of r.v.'s, let*
$$Y_n = \sup_{i,j\geq n} \left|X_i - X_j\right|, Z_n = \sup_k |X_{n+k} - X_n|$$

TFAE:

1. *X_n converges a.s.*
2. *$\lim_{m,n\to\infty} |X_n - X_m| = 0$ a.s.*
3. *$Y_n \to 0$ a.s.*
4. *$Z_n \to 0$ a.s.*

Proposition 3.5. *The follows are sufficient for convergence a.s.:*

1. *For every $\epsilon > 0$*
$$\sum_{n\to\infty} \mathbb{P}\{|X_n - X| > \epsilon\} < \infty$$
2. *There exist a sequence $(\epsilon_i)_{i\in\mathbb{N}^*} \searrow 0$ s.t.*
$$\sum_{n\to\infty} \mathbb{P}\{|X_n - X| > \epsilon\} < \infty$$
3. *There exist a sequence $(\epsilon_i)_{i\in\mathbb{N}^*} > 0$ s.t.*
$$\sum_{n\to\infty} \epsilon_n < \infty, \sum_{n\to\infty} \mathbb{P}\{|X_{n+1} - X_n| > \epsilon_n\} < \infty$$
4. *For every $\epsilon > 0$,*
$$\liminf_{n\to\infty} \lim_{m\to\infty} \mathbb{P}\{\sup_{k\leq m} |X_{n+k} - X_n| > \epsilon\} = 0$$

Proof.

1. By Borel-Cantelli lemma 3.4, that implies statement 2 in theorem 3.2 and thus implies $X_n \to X$ a.s.
2. As $\epsilon_n \to 0$, the equality also holds for all ϵ as one can always find a less ϵ from $(\epsilon_i)_{i\in\mathbb{N}^*}$ and thus 2 $\implies$ 1.
3. By Borel-Cantelli lemma and proposition 3.2.
4. Write Z_n for $\sup_{k\geq 1} |X_{n+k} - X_n|$, by remark 5 in definition 3.2:
$$0 = \liminf_{n\to\infty} \mathbb{P}\{Z_n > \epsilon\} \geq \mathbb{P}(\liminf_{n\to\infty}\{Z_n > \epsilon\}) = 0$$
thus $\{Z_n > \epsilon \text{ ult.}\}$ has probability 0 and thus $\{Z_n > \epsilon \text{ i.o.}\}$ is so, then (Z_n) is converges and so is (X_n) as theorem 3.2.3 and theorem 3.3.

□

3.3 Convergence in Probability

Throughout, $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space and $(X_i)_{i\in\mathbb{N}^*}$ is a subsequence of real-valued r.v.'s.

Definition 3.3. Sequence (X_i) is said to be converge to X **in probability** if, for every $\epsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}\{|X_n - X| > \epsilon\} = 0$$

and denoted as $X_n \xrightarrow{p} X$.

The following relates a.s. convergence and convergence in probability:

Theorem 3.4. *1. $X_n \to X$ a.s. $\implies X_n \xrightarrow{p} X$*

2. If $X_n \xrightarrow{p} X$, then there exists a subsequence that converges to X a.s.

3. If every subsequence has further subsequence that converges to X a.s., then $X_n \xrightarrow{p} X$.

Proof. 1. By theorem 3.2.2, $i_\epsilon \circ |X_n - X| \to 0$ a.s., then $\mathbb{E}\, i_\epsilon \circ |X_n - X| \to 0$ by DCT 1.12 and thus $\mathbb{P}\{|X_n - X| > \epsilon\} \to 0$.

2. As $\lim_{n\to\infty} \mathbb{P}\{|X_n - X| > \epsilon\} = 0$, we can select a subsequence such that

$$\sum_{i\in\mathbb{N}^*} \mathbb{P}\{\left|X_{n_i} - X\right| > \epsilon\} < \infty$$

then claim follows as proposition 3.5.1.

3. Suppose $p_n = \mathbb{P}\{|X_n - X| > \epsilon\}$ as a sequence and $N \subset \mathbb{N}^*$ is subsequence along which the sequence converges to, say, p. By assumption, there is a further subsequence $N' \subset N$ such that $(p_i)_{i\in N'} \to 0$ and that implies $p = 0$. By proposition 3.3, $p_n \to 0 \iff X_n \xrightarrow{p} X$.

□

3.3.1 Convergence and continuous

As an application of above theorem, we have

Proposition 3.6. *Let $f : \mathbb{R} \to \mathbb{R}$ be continuous. Then $f(X_n) \xrightarrow{p} f(X)$ provided $X_n \xrightarrow{p} X$.*

Proof. For every subsequence $N \subset \mathbb{N}^*$, $X_n \xrightarrow{p} X$ along which and theorem 3.4.2 implies $N' \subset N$ exists such that $X_n \to X$ a.s. along and thus $f(X_n) \to X$ a.s. along N'. It follows that $f(X_n) \xrightarrow{p} f(X)$ by theorem 3.4.3.

□

That implies convergence in probability is preserved under arithmetical operations, i.e., if $X_n \xrightarrow{p} X$ and $Y_n \xrightarrow{p} Y$, we have

$$X_n + Y_n \xrightarrow{p} X + Y, X_n - Y_n \xrightarrow{p} X - Y$$
$$X_n Y_n \xrightarrow{p} XY, \frac{X_n}{Y_n} \xrightarrow{p} \frac{X}{Y}$$

where the last equality holds when Y and Y_n are non-zero a.s.

3.3.2 Metric for convergence in probability

For real-valued r.v.'s X and Y, define

$$d(X, Y) = \mathbb{E}(|X - Y| \wedge 1)$$

one can check d is a metric(except we treat X and Y are the same when $X = Y$ a.s.).

The following shows that d can induced convergence in probability.

Proposition 3.7.

$$\lim_{n\to\infty} d(X_n, X) = 0 \iff X_n \xrightarrow{p} X$$

Proof. Note for $\epsilon \in (0, 1)$ and $x \geq 0$:

$$\epsilon i_\epsilon(x) \leq x \wedge 1 \leq \epsilon + i_\epsilon(x)$$

replace x with $|X_n - X|$ and take expectations:

$$\epsilon\, \mathbb{E}\, i_\epsilon \circ |X_n - X| \leq d(X_n, X) \leq \epsilon + \mathbb{E}\, i_\epsilon \circ |X_n - X|$$

thus $\mathbb{E}\, i_\epsilon \circ |X_n - X| = \mathbb{P}\{|X_n - X| > \epsilon\} \to 0$ iff $d(X_n, X) \to 0$ as ϵ can be taken arbitrary small.

□

3.3.3 Cauchy criterion for convergence in probability

Theorem 3.5. *Sequence $(X_i)_{i\in\mathbb{N}^*}$ converges in probability iff for every $\epsilon > 0$,*

$$\lim_{m,n\to\infty} \mathbb{P}\{|X_m - X_n| > \epsilon\} = 0$$

3.4 Convergence in L^p

Definition 3.4. A sequence $(X_i)_{i\in\mathbb{N}^*}$ is said to converges to X in L^p iff $(X_i) \subset L^p$ and $X \in L^p$ and $\|X_n - X\|_p \to 0$.

Converges in L^p also implies convergence in probability by Chebyshev's inequality 2.10 and taking g corresponding to the power:

$$\mathbb{P}\{|X_n - X| > \epsilon\} \leq (\frac{1}{\epsilon})^p\, \mathbb{E}\, |X_n - X|^p \to 0$$

3.4.1 Convergence, Cauchy, uniform integrability

Theorem 3.6. *Suppose $(X_i)_{i\in\mathbb{N}^*}$ taking values in $\mathbb{R}$ and $p \geq 1$, TFAE:*

1. *It converges in L^p.*
2. *It's cauchy in L^p, i.e.:*

$$\lim_{m,n\to\infty} \mathbb{E}\,|X_m - X_n|^p = 0$$

3. *It converges in probability and (X_n^p) is uniformly integrable.*

Proof. $a \implies b$. By the triangle inequality:

$$\|X_m - X_n\|_p \leq \|X_m - X\|_p + \|X - X_n\|_p \to 0$$

and thus $\|X_m - X_n\|_p \to 0$.

$b \implies c$ By Chebyshev's-inequality again and theorem 3.5, it converges in probability. By theorem 2.11, it's sufficient to show that $\forall \epsilon > 0, \exists \delta > 0 \ni \forall A \in \mathcal{F}$,

$$\mathbb{P}(A) \leq \delta \implies \sup_n \mathbb{E}\,|X_n^p|\,\mathbf{1}_A \leq \epsilon$$

and (X_n^p) is L^1 bounded.

The cauchy yields $\mathbb{E}\,|X_m - X_n|^p \leq \epsilon$ for sufficient large $m, n \geq k \gg 1$, thus, for every event $A \in \mathcal{F}$:

$$\mathbb{E}\,|X_n^p| = \mathbb{E}\,|X_n|^p \leq 2^{p-1}(\mathbb{E}\,|X_n - X_k|^p + \mathbb{E}\,|X_k|^p) \leq 2^{p-1}(\epsilon + \mathbb{E}\,|X_k^p|)$$

thus

$$\sup_n \mathbb{E}\,|X_n^p|\,\mathbf{1}_A \leq 2^{p-1}\epsilon + 2^{p-1}\sup_{n\leq k} \mathbb{E}\,|X_n^p|\,\mathbf{1}_A$$

In view of remark 2 in definition 2.6, $\{X_n : n \leq k\}$ is uniformly integrability and thus the right side can be arbitrary small if $\mathbb{P}(A)$ is sufficient small and thus the left side. It follows that (X_n^p) is L^1 bounded by taking $A = \Omega$ and claim follows.

$3 \implies 1$. Let X be the limit. Then $X_n^p \xrightarrow{p} X^p$ by proposition 3.6. By theorem 3.4.2, there is a subsequence $X_n'^p \to X^p$ a.s. then Fatou's lemma 1.7 yields

$$\mathbb{E}\,|X^p| = \mathbb{E}\liminf_n |X_n'^p| \leq \liminf_n \mathbb{E}\,|X_n'^p| \leq \sup_n \mathbb{E}\,|X_n^p| < \infty$$

thus $X^p \in L^1$. Let $F_n = \{|X_n - X|^p > \epsilon\}$, note

$$\mathbb{E}\,|X_n - X|^p \leq \epsilon + \mathbb{E}\,|X_n - X|^p\,\mathbf{1}_{F_n}$$

As X^p is integrable and X_n^p is uniformly so, $(X_n - X)^p$ is uniformly integrable and thus the right side can be arbitrary small if $\mathbb{P}(F_n)$ can be also arbitrary small. It follows that $\mathbb{E}\,|X_n - X|^p \to 0$.

□

The following is a variation of the main results when $p = 1$

Theorem 3.7. *If $X_n \xrightarrow{p} X$, TFAE:*

1. $X_n \to X$ *in* L^1.
2. (X_n) *is uniformly integrable.*
3. $(X_n) \cup \{X\} \subset L^1$ *and* $\mathbb{E}\,|X_n| \to \mathbb{E}\,|X|$.

3.4.2 Convergence of expectations, weak convergence in L^1.

Note convergence in L^1 allows taking limits inside: $X_n \to X$ in L^1 implies $\mathbb{E}\,X_n = \mathbb{E}\,X$:

Definition 3.5. A sequence $(X_n) \subset L^1$ is said to be **converge weakly** in L^1 to X if

$$\lim_{n\to\infty} \mathbb{E}\,X_n Y = \mathbb{E}\,XY$$

holds for all $Y \in \mathcal{F}_b$.

Remark. Where the bounded condition can be replaced by a.s. bounded, then Y can be taken in L^∞. Such convergence induce a topology on L^1 and denoted by $\sigma(L^1, L^\infty)$.

Proposition 3.8. *If $X_n \to X$ in L^1, then it's converge weakly in L^1.*

Proof. Supposing that $|Y| \le b$, if $X_n \to X$ in L^1, then

$$|\mathbb{E}\,X_n Y - \mathbb{E}\,XY| \le \mathbb{E}\,|X_n Y - XY| \le b\,\mathbb{E}\,|X_n - X| \to 0$$

□

Weak convergence implies a deep result:

Proposition 3.9. *Sequence $(X_i)_{i\in\mathbb{N}^*}$ is uniformly integrable iff it's every subsequence has a further subsequence that converges weakly in L^1.*

3.5 Weak Convergence

Throughout, $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space and $(X_i)_{i\in\mathbb{N}^*}$ is a sequence of real-valued r.v.'s with corresponding distribution $(\mu_i)_{i\in\mathbb{N}^*}$, quantile $(q_i)_{i\in\mathbb{N}^*}$ and d.f. $(c_i)_{i\in\mathbb{N}^*}$. See distribution and quantile et seq..

Definition 3.6. Sequence $(\mu_i)_{i\in\mathbb{N}^*}$ is said to be converge weakly to μ iff for any $f \in \mathbb{C}_b$

$$\lim_{n\to\infty} \int f d\mu_n = \int f d\mu$$

Sequence (X_n) is said to converge in **distribution** to X if $\mu_n \to \mu$ weakly, that is

$$\mathbb{E} f(X_n) \to \mathbb{E} f(X)$$

for every $f \in \mathbb{C}_b$ and denoted as $X_n \xrightarrow{d} X$

Remark. $\xrightarrow{p} \implies \xrightarrow{d}$ as every subsequence has a further subsequence converges to X a.s. by theorem 3.4, then $\mathbb{E} f \circ X_n \to \mathbb{E} f \circ X$ along this further subsequence, it follows that $\mathbb{E} f \circ X_n \to \mathbb{E} f \circ X$ by proposition 3.4.

When X is degenerate, i.e., $X = x_0$ a.s., then $\xrightarrow{d} \implies \xrightarrow{p}$ by taking $f(x) = |x - x_0| \wedge 1$ and applying proposition 3.7.

3.5.1 Characterization theorem

Theorem 3.8. *Suppose A is borel set in $\mathbb{R}$, TFAE:*

1. *$\mu_n \to \mu$ weakly.*
2. *$\limsup \mu_n(A) \le \mu(A)$ for every A is closed.*
3. *$\liminf \mu_n(A) \ge \mu(A)$ for every A is open.*
4. *$\mu_n(A) \to \mu(A)$ for every A with $\mu(\partial A) = 0$*

Proof. 1 $\implies$ 2. Suppose A is closed, let $A_\epsilon = \{x : d(x, A) < \epsilon\}$. Then $A_\epsilon \to \overline{A} = A$ as $\epsilon \to 0$ and thus $\mu(A_\epsilon) \searrow \mu(A)$ as $\epsilon \searrow 0$. For any ϵ, define $f(x) = (1 - \frac{d(x,A)}{\epsilon}) \vee 0$, clearly $f \in \mathbb{C}_b$ and $\mathbf{1}_A \le f \le \mathbf{1}_{A_\epsilon}$. Hence

$$\mu_n(A) \le \mu_n f \to \mu f \le \mu(A_\epsilon) \searrow \mu(A)$$

It follows that $\limsup \mu_n(A) \le \mu(A)$.

2 $\iff$ 3. Suppose A is open, then we have

$$\liminf \mu_n(A) = \liminf(1 - \mu_n(A^c)) = 1 - \limsup \mu_n(A^c) \ge 1 - \mu(A^c) = \mu(A)$$

similarly, we can show that 3 $\implies$ 2.

3 $\implies$ 4, By 3 and 2, we have

$$\mu(\overline{A}) \ge \limsup \mu_n(\overline{A}) \ge \limsup \mu_n(A) \ge \liminf \mu_n(A) \ge \liminf \mu_n(A^\circ) \ge \mu(A^\circ)$$

note $\mu(\partial A) = 0 \iff \mu(\overline{A}) = \mu(A^\circ)$, thus the inequalities becomes equalities and thus $\lim_{n\to\infty} \mu_n(A) = \mu(A)$

4 $\implies$ 1. Note borel indicator can approximate any $f \in \mathbb{C}_b$.

□

As borel is π system, weak limits is unique

3.5.2 Convergence of quantiles and distribution functions

Theorem 3.9. *TFAE:*

1. $\mu_n \to \mu$ *weakly.*
2. $c_n(x) \to c(x)$ *for every continuity point* x *of* c.
3. $q_n(u) \to q(u)$ *for every continuity point* u *of* q.

Proof. 1 $\implies$ 2. Let x be a continuity point of c, then $\mu\{x\} = \mathbb{P}\{X = x\} = 0$. Note $\partial(-\infty, x] = \{x\}$, it follows that

$$c_n(x) = \mu_n(-\infty, x] \to \mu(-\infty, x] = c(x)$$

2 $\implies$ 3. Let u be continuity point of q and $x = q(u)$, for any ϵ, pick $y \in (x - \epsilon, x)$ and z in $(x, x + \epsilon)$ such that they are continuity points for c, we can do so as discontinuity points are countable. As q is continues at u, c is not flat at level u and thus $c(y) < u < c(z)$. As $c_n(y) \to c(y)$, we have

$$c_n(y) < u \implies q_n(u) > y > x - \epsilon$$

for tail n and thus $\liminf q_n(u) > x - \epsilon$. Similarly, we have $\lim_{n\to\infty} q_n(u) < x + \epsilon$. Since ϵ can be arbitrary small, we have $q_n(u) \to x = q(u)$.

3 $\implies$ 1. Note discontinuities are at most countable and thus $q_n \to q$ a.s., it follows that for $f \in \mathbb{C}_b$, $f \circ q_n \to f \circ q$ a.s. and hence

$$\mu_n f = \lambda(f \circ q_n) \to \lambda(f \circ q) = \mu f$$

by DCT 1.12. That is, (μ_n) converges to μ weakly.

□

3.5.3 Almost sure representations of weak convergence

Theorem 3.10. *The sequence* (μ_n) *converges weakly to* μ *iff there exist corresponding r.v.'s* $(Y_n), Y$ *on some probability space* $(\Omega', \mathcal{F}', \mathbb{P}')$ *and* $Y_n \to Y$ *a.s. on* $\mathbb{P}'$.

Proof. $\impliedby$ is obvious as $\xrightarrow{a.s.} \implies \xrightarrow{d}$.

$\implies$. Let $q_n = Y_n$, the distribution of Y_n is $\mathbb{P}' \circ q_n^{-1} = \lambda \circ q_n^{-1} = \mu_n$ as desired and live in probability space $((0,1), \mathcal{B}, \lambda)$. By theorem 3.9, $Y_n \to Y$. a.s.

□

This theorem is quite useful in the case of only the distribution are matter.

Proposition 3.10. *Suppose* $X_n \xrightarrow{d} X$, *TFAE:*

1. (X_n) *is uniformly integrable.*
2. $(X_n) \cup \{X\} \in L^1$ *and* $\mathbb{E}|X_n| \to \mathbb{E}|X|$.

If $X_n \xrightarrow{d} X$ and $\mathbb{P}\{X_n \neq Y_n\} \to 0$, then $Y_n \xrightarrow{X}$ in distribution.

Proof. By theorem 3.10, this is immediate from theorem 3.7. □

Note absence here of one statement in theorem 3.7. This is because L^1 convergence concern the joint distribution to determine $X_i - X$.

3.5.4 Construction of convergence in distribution

Proposition 3.11. *Suppose $X_n \xrightarrow{d} X$, $Y_n \xrightarrow{d} Y$ and they are independent, $a, b \in \mathbb{R}$:*

1. $\mathbb{E}|X| \leq \liminf \mathbb{E}|X_n|$.
2. $X_n + Y_n \xrightarrow{d} X + Y$.
3. $a + bX_n \xrightarrow{d} a + bX$.
4. *Suppose Y is degenerate to y, then $X_n Y_n \xrightarrow{d} Xy$ and $X_n + Y_n \xrightarrow{d} X_n + y$.*
5. *Suppose $\mathbb{P}\{X_n \neq Y_n\} \to 0$, then $Y_n \xrightarrow{d} X$.*

Proof.

1. By theorem 3.10, consider $U_n \xrightarrow{a.s.} U$ share the same distribution with X_n and X and thus share the same expansion, the claim follows by Fatou's lemma 1.7.
2. Follows from Continuous mapping theorem.
3. Follows from Continuous mapping theorem.
4. Follows from Continuous mapping theorem.
5. Let c_n, c, c'_n by $d.f.$ corresponding to X_n, X, Y_n, note that

$$c_n(x) = \mathbb{P}\{X_n \leq x\} \leq \mathbb{P}\{X_n \neq Y_n\} + \mathbb{P}\{Y_n \leq x\} \leq \mathbb{P}\{X_n \neq Y_n\} + c'_n(x)$$

similarly results holds for $c'_n(x)$, thus

$$c_n(x) - \mathbb{P}\{X_n \neq Y_n\} \leq c'_n(x) \leq c_n(x) + \mathbb{P}\{X_n \neq Y_n\}$$

Then claim follows by assumption and $c_n \to c$.

□

3.5.5 Tightness and Prohorov's theorem

Definition 3.7. Sequence $(\mu_i)_{i\in\mathbb{N}^*}$ is said to be **tight** if for every $\epsilon > 0$, there is a compact K such that $1 - \mu_n(K) = \mu_n(K^c) < \epsilon$ for all n.

Theorem 3.11 (Prohorov's theorem)**.** *If μ_n is tight then every subsequence has a further subsequence converges weakly.*

Proof. By Helly's theorem 3.1, for each subsequence $N \subset \mathbb{N}$, there is a further subsequence converges to some d.f. c pointwise on the continuity set of c. In view of theorem 3.9, it's sufficient to show that the corresponding μ of c is a probability measure, i.e., $c(\infty) = 1$ and $c(-\infty) = 0$. For any ϵ, as (μ_n) is tight, there is a compact $[a, b]$ interval such that $\mu_n[a, b] > 1 - \epsilon$. Select continuity x of c from $(-\infty, a)$ and y from (b, ∞). Then

$$\begin{aligned} c_n(x) &= \mu_n(-\infty, x) \leq \mu_n(-\infty, a) \leq \mu_n[a, b]^c < \epsilon \\ c_n(y) &= \mu_n(-\infty, y) \geq \mu_n[a, b] > 1 = \epsilon \end{aligned}$$

That implies $c(-\infty) \leq c(x) < \epsilon$ and $c(\infty) \geq c(y) > 1 - \epsilon$ and the claim follows. □

Remark. That implies $\mu_n \to \mu$ weakly in view of proposition 3.4 if every further subsequence converges to the same μ.

3.5.6 Convergence of ch.f.

Let φ_n be corresponding ch.f. of μ_n, i.e. $\varphi_n(r) = \mathbb{E}\, e^{irx}$, the next theorem connects the convergence of (μ_n) and (φ_n).

Theorem 3.12 (Levy's continuity theorem)**.** *The sequence (μ_n) is weakly converges to a distribution μ iff $\lim_{n\to\infty} \varphi_n(r) \to \varphi(r)$ for every $r \in \mathbb{R}$ and φ is continuous at 0. Moreover, φ is precisely ch.f. of μ.*

Proof. $\Longrightarrow$ is immediate from $\cos(rx)$ and $\sin(rx)$ are both continuous and bounded and hence

$$\varphi_n(r) = \mu_n \cos(rx) + i\mu_n \sin(rx) \to \mu \cos(rx) + i\mu \sin(rx) = \varphi(r)$$

and the continuity of 0 follows from uniformly continuity of φ.

$\Longleftarrow$. Let μ be corresponding distribution of φ, for any $\epsilon > 0$, note

$$\begin{aligned} \frac{1}{2\epsilon} \int_{(-\epsilon,\epsilon)} \varphi(t)dt &= \frac{1}{2\epsilon} \int_{(-\epsilon,\epsilon)} \int e^{itx} d\mu dt \\ &= \frac{1}{2\epsilon} \int \int_{(-\epsilon,\epsilon)} e^{itx} dt d\mu \\ &= \frac{1}{2\epsilon} \int \int_{(-\epsilon,\epsilon)} \cos tx dt d\mu \\ &= \int \frac{\sin \epsilon x}{\epsilon x} d\mu = \mu(\frac{\sin \epsilon x}{\epsilon x}) \end{aligned}$$

Then we show that (μ_n) is tight, for any $M > 0$,

$$\begin{aligned}
\left|\frac{1}{2\epsilon}\int_{(-\epsilon,\epsilon)} \varphi_n(t)dt\right| &\le \mu_n \left|\frac{\sin \epsilon x}{\epsilon x}\right| \\
&= \mu_n \left|\frac{\sin \epsilon x}{\epsilon x}\right| \mathbf{1}_{[-M,M]} + \mu_n \left|\frac{\sin \epsilon x}{\epsilon x}\right| \mathbf{1}_{[-M,M]^c} \\
&\le \mu_n \mathbf{1}_{[-M,M]} + \mu_n \left|\frac{1}{\epsilon x}\right| \mathbf{1}_{[-M,M]^c} \\
&\le \mu_n[-M,M] + \frac{1}{\epsilon M}\mu_n([-M,M]^c) \\
&\le \mu_n[-M,M] + \frac{1}{\epsilon M}
\end{aligned}$$

Let $n \to \infty$, by DCT 1.12 we have

$$\inf \mu_n[-M,M] + \frac{1}{\epsilon M} \ge \frac{1}{2\epsilon}\int_{(-\epsilon,\epsilon)} d\varphi$$

Since φ is continuous at 0, in view of Mean Value Theorem, for any $\epsilon > 0$, we can find δ such that

$$\frac{1}{2\delta}\int_{(-\delta,\delta)} d\varphi > \varphi(0) - \epsilon = 1 - \epsilon$$

and we can find M such that $\frac{1}{\delta M} < \epsilon$ clearly, thus

$$\inf \mu_n[-M,M] \ge 1 - 2\epsilon$$

and thus (μ_n) is tight. Then claim follows as remark in theorem 3.11.

□

3.6 Laws of Large Numbers

Throughout, $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space and $(X_i)_{i\in\mathbb{N}^*}$ is a sequence of real-valued r.v.'s, and for $n \ge 1$, define

$$S_n = \sum_{i=1}^{n} X_i, \overline{X}_n = \frac{S_n}{n}$$

Now we are ready for the important results in classical probability theory. The statement about convergence in probability is called **weak law of large numbers** and the one about *a.s.* convergence is called **strong law of large numbers**.

We start with finite case.

Theorem 3.13. *Suppose (X_n) are pairwise independent and identical distributed with finite mean a and finite variation b. Then $(\overline{X}_n)$ converges to a in L^2(and thus in probability) and a.s.*

Proof. The L^2 convergence follows by noting

$$\mathbb{E}\left|\overline{X}_n - a\right|^2 = \operatorname{Var}\overline{X}_n = \frac{b}{n} \to 0$$

For the $a.s.$ convergence, WLOG, we may assume $X_n \geq 0$ for all n. Let $N = (n_i)$ be subsequence of $\mathbb{N}$ by $n_i = i^2$, by Chebyshev's inequality 2.10.

$$\epsilon^2 \sum_{n\in N} \mathbb{P}\{\left|\overline{X}_n - a\right| > \epsilon\} \leq \sum_{n\in N} \mathbb{E}\left|\overline{X}_n - a\right|^2 = \sum_{n\in N} \frac{b}{n} = b\sum_{i\in\mathbb{N}^*} \frac{1}{i^2} < \infty$$

Thus $\overline{X}_n \to a$ $a.s.$ along N by proposition 3.5. Let Ω_0 be the set witness the $a.s.$ convergence, for $\omega \in \Omega_0$, by lemma 3.1 where $r = 1$, we have:

$$a \leq \liminf \overline{X}_n(\omega) \leq \limsup \overline{X}_n(\omega) \leq a$$

The inequalities should be equalities, which completes the proof. □

3.6.1 Strong law of large numbers

In the preceding lemma we assume (X_n) have finite variation and variation, now we remove them.

Proposition 3.12. *Suppose $X_n \geq 0$ are pairwise independent and identically distributed with $\mathbb{E}\,X_i = \infty$, then $\overline{X}_n \to \infty$ a.s.*

Proof. Fix $b \geq 0$ and let $Y_n = X_n \wedge b$, then theorem 3.13 applies to (Y_n) and $\overline{Y}_n \to \mathbb{E}(X \wedge b)$ $a.s.$ Since $X_n \geq Y_n$ for all n,

$$\liminf \overline{X}_n \geq \lim_{n\to\infty} \overline{Y}_n = \mathbb{E}(X \wedge b)$$

holds $a.s.$ Then claim follows as $\lim_{b\to\infty} \mathbb{E}(X \wedge b) = \mathbb{E}\,X = \infty$ by MCT 1.10. □

Now we are ready to give Etemadi's proof:

Theorem 3.14 (Strong law of large numbers)**.** *Suppose X_n are pairwise independent and identical distributed with X, then $\overline{X}_n \to \mathbb{E}\,X$ a.s. if $\mathbb{E}\,X$ exist.*

Proof. WLOG, assume $X_n \geq 3$(We can do so by replacing $X_n \leftarrow X_n + 3$) and preceding proof guarantee we can assume $3 \leq X_n < \infty$ as $\mathbb{E}\,X < \infty$.

Step 1 Let

$$Y_n = X_n \mathbf{1}_{X_n<n}, T_n = \sum_{i=1}^{n} Y_i, \overline{Y}_n = \frac{T_n}{n}$$

Clearly, Y_n is bounded and do not differ from X_n much since

$$\sum_n \mathbb{P}\{X_n \neq Y_n\} = \sum_n \mathbb{P}\{X_n \geq n\} = \sum_n \mathbb{P}\{X \geq n\} \leq \int_0^\infty \mathbb{P}\{X \geq t\}dt = \mathbb{E}\,X < \infty$$

which implies, through Borel-Cantelli lemma 3.4, $X_n = Y_n$ *a.s.* for all but finitely many n, therefore it's sufficient to show that $\overline{Y}_n \to \mathbb{E}\, X$ *a.s.*

Step 2 (Y_n) remain pairwise independent and hence:

$$\mathbb{E}\, T_n = \sum_{i=1}^{n} \mathbb{E}\, X \mathbf{1}_{X<i} = \mathbb{E}\, X \sum_{i>X} \delta_i[1,n]$$

$$\operatorname{Var} T_n = \sum_{i=1}^{n} \operatorname{Var} Y_i \le \sum_{i=1}^{n} \mathbb{E}\, Y_i^2 = \mathbb{E}\, X^2 \sum_{i>X} \delta_i[1,n]$$

where δ_i is Dirac sitting at i as usual. Let $Z_n = \sum_{i>X} \delta_i[1,n]$, note that it's the number of integers in $(X, n]$, $(\frac{XZ_n}{n})$ is dominated by X and converges to X, by DCT 1.12, we have

$$\mathbb{E}\, \overline{Y}_n = \mathbb{E}\, \frac{XZ_n}{n} \to \mathbb{E}\, X$$

Step 3 Then we find a subsequence $N \subset \mathbb{N}^*$ for which $\overline{Y_n} \to \mathbb{E}\, X$ *a.s.* across it and it's equivalent to $\overline{Y}_n - \mathbb{E}\, \overline{Y}_n \to 0$ *a.s.* across N.

Let $N = (n_i) \subset \mathbb{N}^*$ where $n_i = \lceil e^{ai} \rceil$ for some fixed a. In view of proposition 3.5, it's sufficient to show that $\forall \epsilon > 0$

$$s = \sum_{n \in N} \mathbb{P}\{\left|\overline{Y}_n - \mathbb{E}\, \overline{Y}_n\right| > \epsilon\} < \infty$$

by Chebyshev's inequality:

$$\begin{aligned}
\epsilon^2 s &\le \sum_{n\in N} \operatorname{Var} \overline{Y}_n = \sum_{n \in N} \frac{1}{n^2}\, \mathbb{E}\, X^2 Z_n \\
&= \mathbb{E}\, X^2 \sum_{n\in N} \frac{1}{n^2} Z_n = \mathbb{E}\, X^2 \sum_{i>X} \sum_{n \in N} \frac{1}{n^2} \delta_i[1,n] \\
&\le \mathbb{E}\, X^2 \sum_{i>X} \sum_{\{n \in N : n \ge i\}} \frac{1}{n^2} \\
&= \mathbb{E}\, X^2 \sum_{i>X} \sum_{k \ge m_i} \frac{1}{n_k^2}
\end{aligned}$$

where $m_i = \inf_{\mathbb{N}}\{j : n_j \ge i\}$, note

$$e^{am_i} + 1 > n_{m_i} \ge i \implies e^{am_i} > i - 1$$

thus

$$\begin{aligned}
\sum_{k \ge m_i} \frac{1}{n_k^2} \le \sum_{k \ge m_i} \exp -2ak &= \sum_{j=1}^{\infty} \exp -2a(j + m_i) \\
&= (\sum_{j=1}^{\infty} \exp -2aj) \exp -2am_i \\
&= c \exp -2am_i \le c \frac{1}{(i-1)^2}
\end{aligned}$$

Where we denoted $c = \frac{1}{e^{2a}-1} < \infty$. Then

$$\sum_{i>X} \frac{1}{(i-1)^2} \leq \int_{X-2}^{\infty} \frac{1}{x^2} dx = \frac{1}{X-2}$$

Thus

$$\epsilon^2 s \leq c\,\mathbb{E}\,\frac{X^2}{X-2} \leq c\,\mathbb{E}(X+6) < \infty$$

as $X \geq 3$. Then claim follows as $s < \infty$.

Step 4 Similar to the proof in finite case 3.13, by lemma 3.1 where $r = e^a$:

$$e^{-a}\,\mathbb{E}\,X \leq \liminf \overline{Y}_n(\omega) \leq \limsup \overline{Y}_n(\omega) \leq e^a\,\mathbb{E}\,X$$

which completes the proof by letting $a \to 0$.

□

3.6.2 Weak law of large numbers

In the classical LLN, we completes the proof by $\operatorname{Var} S_n = \sum \operatorname{Var} X_i$ and $\operatorname{Var} \frac{S_n}{n} \to 0$. Both can be ensured by some weaker conditions.

Theorem 3.15. *Suppose that X_n are pairwise uncorrelated and $\sum$ Var $\frac{X_n}{b_n}$ converges for some diverge $b_n > 0$. Then $\frac{S_n - \mathbb{E}\,S_n}{b_n} \to 0$ in L^2.*

Proof. Uncorrelatedness implies

$$\mathbb{E}\left|\frac{S_n - \mathbb{E}\,S_n}{b_n}\right|^2 = \operatorname{Var} \frac{S_n}{b_n} = \frac{1}{b_n^2} \sum \operatorname{Var} X_i$$

then claim follows as lemma 3.3.

□

3.7 Convergence of Series

Now we focus on the *a.s.* convergence of series $S_n = \sum_{i \leq n} X_i$. All the results rests on independence of (X_n), thus we can use Kolmogorov's 0-1 law and thus the convergence of the series has probability 0 or 1.

3.7.1 Inequalities for maxima

Suppose $\mathbb{E}\,X_i = 0$ for each of (X_n), then Chebyshev's inequality 2.10 yields:

$$\epsilon^2\,\mathbb{P}\{|S_n| > \epsilon\} \leq \operatorname{Var} S_n = \mathbb{E}\,S_n^2$$

We can improve it when X_n are independent:

Theorem 3.16 (Kolmogorov's inequality)**.** *Suppose that X_n are independent and have mean 0, then $\forall \epsilon \in (0, \infty)$,*

$$\epsilon^2\, \mathbb{P}\{\max_{k \le n} |S_k| > \epsilon\} \le Var\, S_n$$

Proof. Fix $\epsilon > 0$ and $n \ge 1$, define $N(\omega) = \inf\{k \ge 1 : |S_k(\omega)| > \epsilon\}$, one can check that it's a *r.v.* by noting that $\mathbf{1}_{N=k}$ is a function of $(X_i)_{1 \le i \le k}$. Consequently, for $k < n$, $U = S_k \mathbf{1}_{N=k}$ and $V = S_n - S_k$ are functions of independent $(X_i)_1^k$ and $(X_i)_{k+1}^n$. And thus $\mathbb{E}\, UV = \mathbb{E}\, U\, \mathbb{E}\, V = 0$ as $\mathbb{E}\, V = 0 \iff \mathbb{E}\, X_i = 0$. Hence, for $k \le n$:

$$\mathbb{E}\, S_k (S_n - S_k) \mathbf{1}_{N=k} = 0$$

Note $S_n^2 \ge S_k^2 + 2 S_k (S_n - S_k)$ and $|S_k|^2 > \epsilon^2$ on the event $\{N = k\}$. Thus

$$\mathbb{E}\, S_n^2 \mathbf{1}_{N=k} \ge \epsilon^2\, \mathbb{E}\, \mathbf{1}_{N=k} + 2\, \mathbb{E}\, S_k (S_n - S_k) \mathbf{1}_{N=k} = \epsilon^2\, \mathbb{P}\{N = k\}$$

Summing both sides:

$$\epsilon^2\, \mathbb{P}\{N \le n\} \le \mathbb{E}\, S_n^2 \mathbf{1}_{N \le n} \le \mathbb{E}\, S_n^2 = \mathrm{Var}\, S_n$$

Then claim follows from $\{N \le n\} = \{\max_{k \le n} |S_k| > a\}$.

□

The assumption of independence for the (X_n) will be relaxed later by martingaling. For the present, the following is an estimate going in the opposite direction.

Lemma 3.5. *Suppose (X_n) are independent with zero mean and dominated by some constant M, then $\forall \epsilon > 0$,*

$$\mathbb{P}\{\max_{k \le n} |S_k| > \epsilon\} \ge 1 - \frac{(M + \epsilon)^2}{Var\, S_n}$$

Proof. Fix n and ϵ and define N as preceding proof. Now the claim is

$$\mathbb{P}\{N > n\}\, \mathrm{Var}\, S_n \le (M + \epsilon)^2$$

For $k \le n$, note $S_n^2 = S_k^2 + 2 S_k (S_n - S_k) + (S_n - S_k)^2$ and $|S_k| = |X_k + S_{k-1}| \le M + \epsilon$ if $N = k$. Thus

$$\begin{aligned}
\mathbb{E}\, S_n^2 \mathbf{1}_{N=k} &\le (M + \epsilon)^2\, \mathbb{E}\, \mathbf{1}_{N=k} + 2\, \mathbb{E}\, S_k (S_n - S_k) \mathbf{1}_{N=k} + \mathbb{E} (S_n - S_k)^2 \mathbf{1}_{N=k} \\
&= (M + \epsilon)^2\, \mathbb{E}\, \mathbf{1}_{N=k} + \mathbb{E} (S_n - S_k)^2 \mathbf{1}_{N=k} \\
&\le (M + \epsilon)^2\, \mathbb{E}\, \mathbf{1}_{N=k} + \mathbb{E}\, S_n^2\, \mathbb{E}\, \mathbf{1}_{N=k}
\end{aligned}$$

Summing over $k \le n$,

$$\mathbb{E}\, S_n^2 \mathbf{1}_{N \le n} \le [(M + \epsilon)^2 + \mathrm{Var}\, S_n]\, \mathbb{P}\{N \le n\}$$

On the other hand, if $N < n$, we have $|S_n| \le \epsilon$ and hence

$$\mathbb{E}\, S_n^2 \mathbf{1}_{N > n} \le \mathbb{E}\, \epsilon^2 \mathbf{1}_{N > n} = \epsilon^2\, \mathbb{P}\{N > n\}$$

Then claim follows by adding them and some rearrangement.

□

3.7.2 Convergence of series and variance

Theorem 3.17. *Suppose (X_n) are independent with zero mean, then $Var S_n = \sum Var X_n$ converges implies $S_n = \sum X_n$ converges a.s.*

Proof. Apply Kolmogorov's inequality 3.16 to $(X_i)_{i=n+1}^{n+m} = S_{n+m} - S_n$, $\forall \epsilon > 0$,

$$\epsilon^2 \, \mathbb{P}\{\max_{k \le m} |S_{n+k} - S_n| > \epsilon\} \le \sum_{i=n+1}^{n+m} \operatorname{Var} X_i$$

By assumption, the right side goes to 0 as $m \to \infty$ and $n \to \infty$, then (S_n) converges in view of proposition 3.5.4.

□

The following is a partial converse:

Proposition 3.13. *Suppose (X_n) are bounded and independent. If $\sum(X_n - a_n)$ converges a.s. for some $(a_n) \subset \mathbb{R}$, then $\sum Var X_i < \infty$*

Proof. **Step 1** Start with extra condition that $a_n = 0$ and $\mathbb{E}\, X_n = 0$ for all n. Let b be a bound for X_n, note

$$Z_m = \sup_k |S_{m+k} - S_m| = \lim_{n \to \infty} \max_{k \le n} |S_{m+k} - S_m|$$

Thus, for any $\epsilon > 0$, by lemma 3.5:

$$\mathbb{P}\{Z_m > \epsilon\} = \lim_{n \to \infty} \mathbb{P}\{\max_{k \le n} |S_{m+k} - S_m| > \epsilon\} \ge 1 - \frac{(\epsilon + b)^2}{\sum_{i=m+1}^{\infty} \operatorname{Var} X_i}$$

Note $Z_m \to 0$ *a.s.* in view of theorem 3.3 and thus $Z_m \xrightarrow{p} 0$, that implies $\sum \operatorname{Var} X_i$ can not diverge.

Step 2 Suppose (Y_n) and (X_n) are independent and share the same the law, then $\sum(Y_n - a_n)$ converges *a.s.* and thus $\sum(X_n - Y_n)$ is so. Apply the result in step 1 to $(X_n - Y_n)$ and claim follows by noting $\operatorname{Var}(X_n - Y_n) = 2 \operatorname{Var} X_n$.

□

3.7.3 Kolmogorov's three series theorem

We have given the necessary and sufficient conditions for the *a.s.* convergence of S_n, then we are ready to combine them. As (X_n) is generally not bounded, define

$$Y_n = X_n \mathbf{1}_{|X_n| \le b}$$

where $b > 0$.

Theorem 3.18. *Suppose (X_n) are independent then S_n is converges a.s. iff so are the following:*

$$\sum \mathbb{P}\{X_n \ne Y_n\}, \sum \mathbb{E}\, Y_n, \sum Var Y_n$$

Proof. $\Longleftarrow$. Clearly, (Y_n) is also independent. Apply theorem 3.17 to $(Y_n - \mathbb{E}\,Y_n)$ then the convergence of $\sum \operatorname{Var} Y_n$ implies $\sum(Y_n - \mathbb{E}\,Y_n)$ converges *a.s.*. That along with convergence of $\sum \mathbb{E}\,Y_n$ implies $\sum Y_n$ converges *a.s.* It follows that $X_n = Y_n$ *a.s.* for all but finite many n by lemma 3.4 and claim follows.

$\Longrightarrow$. Suppose (S_n) converges *a.s.*, let Ω_0 be the set where (S_n) converges. For $\omega \in \Omega_0$, there are at most finitely many n *s.t.* $|X_n(\omega)| > b$, which in turn implies $X = Y$ in Ω_0 for all but finitely many n, *i.e.*, $\mathbb{P}\{X_n \neq Y_n \text{ i.o.}\} = 0$. In view of lemma 3.4, $\sum \mathbb{P}\{X_n \neq Y_n\}$ must converges.

On the other hand, $\sum Y_n$ converges *a.s.*, then $\sum \operatorname{Var} Y_n$ follows by proposition 3.13 and so is $\sum(Y_n - \mathbb{E}\,Y_n)$ by theorem 3.17, that together with convergence of $\sum Y_n$ implies the convergence of $\sum \mathbb{E}\,Y_n$.

□

3.7.4 Application to strong laws

By Kolmogorov's series theorem, we can improve results in theorem 3.15 in independent case.

Proposition 3.14. *Suppose that X_n are independent and $\sum$ Var $\frac{X_n}{b_n}$ converges for some diverge $b_n > 0$. Then $\frac{S_n - \mathbb{E}\,S_n}{b_n} \to 0$ in L^2 and a.s.*

Proof. Apply theorem 3.17 to $(\frac{X_n - \mathbb{E}\,X_n}{b_n})$, we have $\sum \frac{X_n - \mathbb{E}\,X_n}{b_n}$ converges *a.s.*, then claim follows by Kronecker's lemma 3.3.

□

3.8 Central Limits

Start with a generalization of DeMoivre-Laplace theorem:

Theorem 3.19 (DeMoivre-Laplace Theorem)**.** *Let $(X_i)_{i \in \mathbb{N}^*}$ are i.i.d. Bernoulli variables with mean $\mu = p$ and variance $\sigma^2 = p(1-p)$, then*

$$\frac{S_n - n\mu}{\sqrt{n\sigma^2}} \xrightarrow{d} \mathfrak{z}$$

Theorem 3.20 (Lindeberg-Levy Theorem)**.** *Let $(X_i)_{i \in \mathbb{N}^*}$ be i.i.d. with mean μ and variance σ^2, both finite, then*

$$Z_n = \frac{S_n - n\mu}{\sqrt{n\sigma^2}} \xrightarrow{d} \mathfrak{z}$$

Proof. In view of theorem 3.12, the claim is

$$\varphi_{Z_n}(t) \to \varphi_{\mathfrak{z}}(t) = e^{-t^2/2}$$

Let φ denote the $ch.f.$ of $\frac{X_n-\mu}{\sigma}$, then Taylor's theorem yields

$$\begin{aligned}\varphi(t) &= \varphi(0) + \varphi'(0)t + \frac{1}{2}\varphi''(0)t^2(1+h(t)) \\ &= 1 - \frac{1}{2}t^2(1+h(t))\end{aligned}$$

for some h $s.t.$ $\lim_{t\to\infty}|h(t)| = 0$. As (X_n) are independent, note $Z_n = \sum \frac{X_n-\mu}{\sigma}/\sqrt{n}$:

$$\varphi_{Z_n}(t) = \varphi^n(\frac{t}{\sqrt{n}}) = [1 - \frac{r^2/2}{n}(1+h(\frac{r}{\sqrt{n}}))]^n$$

note $\frac{r}{\sqrt{n}} \to 0$ as $n \to \infty$, thus

$$[1 - \frac{r^2/2}{n}(1+h(\frac{r}{\sqrt{n}}))]^n \to (1 - \frac{r^2/2}{n})^n \to e^{-r^2/2}$$

and claim follows.

$\square$

3.8.1 Triangular matrix

Throughout, we shall deal with a infinite random matrix:

$$\mathbf{X} = \begin{bmatrix} X_{11} & X_{12} & X_{13} & \dots \\ X_{21} & X_{22} & X_{23} & \dots \\ X_{31} & X_{32} & X_{33} & \dots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

where each entry is real-valued, for each i, there is an k_i $s.t.$ $\mathbf{X}_{ij} = 0$ for all $j > k_i$ and $k_i \to \infty$. Thus the random matrix is basically triangular. We let Z_i denoted as row sum, $Z_i = \sum_j X_{ij}$ and variables among each row are independent.

3.8.2 Liapunov's Theorem

Following lemma is useful for Liapunov's Theorem.

Lemma 3.6. *Let $(Y_i)_1^k$ be independent and have mean* 0. *Let S be their sum and assume $Var\,S = 1$. Let $f : \mathbb{R} \to \mathbb{R}$ be differentiable thrice and assume that the derivatives are all bounded and continues where f''' are bounded by c. Then*

$$|\mathbb{E}\, f(S) - \mathbb{E}\, f(\mathfrak{Z})| \leq c\sum_{j=1}^{k} \mathbb{E}\left|Y_j\right|^3$$

Proof. Let $(Z_i)_{i\in\mathbb{N}^*}$ be independent with distribution $\mathcal{N}(0, \operatorname{Var} Y_j)$ respectively, then $T = \sum_{i=1}^k Z_i$ has the same distribution with $\mathfrak{Z}$ and the claim is:

$$|\mathbb{E}\, f(S) - \mathbb{E}\, f(T)| \le c \sum_{j=1}^{k} \mathbb{E}\left|Y_j\right|^3$$

The idea is to replace Y_i with Z_i one by one, define

$$V_i = S + \sum_{j<i} (Z_j - Y_j) - Y_i = \sum_{j>i} Y_j + \sum_{j<i} Z_j$$

where V_i is independent with Y_i and Z_i, note $V_k + Z_k = T$ and $V_1 + Y_1 = S$, we have

$$f(S) - f(T) = \sum_{i=1}^{k} [f(V_i + Y_i) - f(V_i + Z_i)]$$

So it's enough to show that

$$|\mathbb{E}\, f(V_i + Y_i) - \mathbb{E}\, f(V_i + Z_i)| \le c\, \mathbb{E}\left|Y_i\right|^3$$

for all $i \le k$. By Taylor's expansion of f at v:

$$f(v + x) = f(v) + f'(v)x + \frac{1}{2} f''(v)x^2 + \frac{1}{6} f'''(\xi)x^3$$

where $|f'''(\xi)| \le c$. Replace v with V and x with V and Z respectively and note Y and Z share the same mean and variance, we have

$$|\mathbb{E}\, f(V_i + Y_i) - \mathbb{E}\, f(V_i + Z_i)| \le \frac{c}{6}\left|\mathbb{E}\, Y_i^3 - \mathbb{E}\, Z_i^3\right| \le \frac{c}{6}(\mathbb{E}\left|Y_i\right|^3 + \mathbb{E}\left|Z_i\right|^3)$$

Direct computation shows that

$$\mathbb{E}\left|Z_i\right|^3 = \sigma^3 \sqrt{\frac{8}{\pi}} \le 2\sigma^3 = 2\|Y_i\|_2^3 \le 2\|Y_i\|_3^3 = 2\,\mathbb{E}\left|Y_i\right|^3$$

and the claim follows.

□

By intuition, some condition on the third moments is sufficient to CLT:

Theorem 3.21 (Liapunov's Theorem)**.** *Suppose that* $\mathbb{E}\,\mathbf{X}_{ij} = 0$*,* $Var\, Z_i = 1$ *and* $\lim_{i\to\infty} \sum_j \mathbb{E}\left|\mathbf{X}_{ij}\right|^3 = 0$*, then,* $Z_i \xrightarrow{d} \mathfrak{Z}$.

Proof. Apply above lemma to $\cos tx$ and $\sin tx$, we get

$$|\mathbb{E}\cos(tZ_i) - \mathbb{E}\cos(t\mathfrak{Z})| \le |t|^3 \sum_j \mathbb{E}\left|\mathbf{X}_{ij}\right|^3$$

$$|i\,\mathbb{E}\sin(tZ_i) - i\,\mathbb{E}\sin(t\mathfrak{Z})| = |\mathbb{E}\sin(tZ_i) - \mathbb{E}\sin(t\mathfrak{Z})| \le |t|^3 \sum_j \mathbb{E}\left|\mathbf{X}_{ij}\right|^3$$

thus

$$\left|\varphi_{Z_i}(t) - \varphi_{\mathfrak{Z}}(t)\right| \leq 2\,|t|^3 \sum_j \mathbb{E}\left|\mathbf{X}_{ij}\right|^3 \to 0$$

It follows that $Z_i \xrightarrow{d} \mathfrak{Z}$ by theorem 3.12.

□

The generalization of Liapunov's theorem is also true.

Corollary 3.1. *Let* $\mathbb{E}\, Z_i = \mu_i$ *and* $Var\, Z_i = \sigma_i^2$*, suppose* $\mu_i \to \mu$ *and* $\sigma_i \to \sigma \neq 0$*, each* $\mathbf{X}_{ij}$ *is bounded by* c_{ij} *and* $\lim_{i\to\infty} \sup_j c_{ij} = 0$*. Then,* $\frac{Z_i - \mu}{\sigma} \xrightarrow{d} \mathfrak{Z}$*.*

Proof. Put $\mathbf{Y}_{ij} = \frac{\mathbf{X}_{ij} - \mathbb{E}\,\mathbf{X}_{ij}}{b_i}$. Note that

$$\left|\mathbf{Y}_{ij}\right| \leq \frac{2c_{ij}}{b_i} \leq \frac{2\sup_j c_{ij}}{b_i} = \epsilon_i$$

and thus $\left|\mathbf{Y}_{ij}\right|^3 \leq \epsilon_i \left|\mathbf{Y}_{ij}\right|^2$. Thus

$$\sum_j \mathbb{E}\left|\mathbf{Y}_{ij}\right|^3 \leq \epsilon_i \sum_j \frac{\operatorname{Var}\mathbf{X}_{ij}}{b_i} = \epsilon_i$$

where $\epsilon_i \to 0$ by assumption and thus Liapunov's theorem 3.21 applies to $\mathbf{Y}_{ij}$ to show:

$$\frac{Z_i - \mu_i}{\sigma_i} = \sum_j \mathbf{Y}_{ij} \xrightarrow{d} \mathfrak{Z}$$

and claim follows by continuity of convergence in distribution.

□

3.8.3 Lindeberg's Theorem

Now we relax the condition on the third moments to **Lindeberg's condition**: $\forall \epsilon > 0$,

$$L_i(\epsilon) = \sum_j \mathbb{E}\,\mathbf{X}_{ij}^2 \mathbf{1}_{|\mathbf{X}_{ij}|>\epsilon} \to 0$$

Select $m(\epsilon)$ *s.t.* $L_n(\epsilon) \leq \epsilon^3$ for all $n \geq m(\epsilon)$, then, choose ϵ_n *s.t.* $m(\epsilon_n) \leq n$ for n large enough, then $\epsilon_n \to 0$ as $n \to \infty$. Then

$$\lim_{n\to\infty} \left(\frac{1}{\epsilon_n}\right)^2 L_n(\epsilon_n) \leq \lim_{n\to\infty} \left(\frac{1}{\epsilon_n}\right)^2 \epsilon_n^3 = \lim_{n\to\infty} \epsilon_n = 0$$

By similar deduction, we have $\frac{1}{\epsilon} L_i(\epsilon) \to 0$ across some ϵ_n.

Theorem 3.22 (Lindeberg's theorem)**.** *WLOG, suppose* $\mathbb{E}\,\mathbf{X}_{ij} = 0$ *and* $Var\, Z_i = 1$ *with Lindeberg's condition, then* $Z_i \xrightarrow{d} \mathfrak{Z}$*.*

Proof. Let $\mathbf{Y}_{ij} = \mathbf{X}_{ij}\mathbf{1}_{|\mathbf{X}_{ij}|\le\epsilon_n}$ and put $S_i = \sum_j \mathbf{Y}_{ij}$, then

$$\mathbb{P}\{Z_i \neq S_i\} \le \sum_j \mathbb{P}\{\mathbf{X}_{ij} \neq \mathbf{Y}_{ij}\} = \sum_j \mathbb{P}\{\left|\mathbf{X}_{ij}\right| > \epsilon_i\} \le \left(\frac{1}{\epsilon_i}\right)^2 L_i(\epsilon_i) \to 0$$

where the last inequality follows by noting that $\epsilon^2\mathbf{1}_{|X|>\epsilon} \le X^2\mathbf{1}_{|X|>\epsilon}$. In view of proposition 3.11, it's sufficient to show that $S_n \xrightarrow{d} \mathfrak{Z}$.

Denoted $\mathbf{X}_{ij} = X$ and $\mathbf{Y}_{ij} = Y$ for short. Since $\mathbb{E}\,X = 0$, $\mathbb{E}\,Y = \mathbb{E}(Y - X) = -\,\mathbb{E}\,X\mathbf{1}_{|X|>\epsilon}$ and thus

$$\left|\mathbb{E}\,Y\right| \le \mathbb{E}\left|X\right|\mathbf{1}_{|X|>\epsilon} \le \frac{1}{\epsilon}\,\mathbb{E}\,X^2\mathbf{1}_{|X|>\epsilon}$$

It follows that

$$\begin{aligned}
\operatorname{Var} Y &= \mathbb{E}\,Y^2 - (\mathbb{E}\,Y)^2 \\
&\ge \mathbb{E}\,X^2\mathbf{1}_{|X|\le\epsilon} - \left(\mathbb{E}\left|X\right|\mathbf{1}_{|X|>\epsilon}\right)^2 \\
&\ge \mathbb{E}\,X^2\mathbf{1}_{|X|\le\epsilon} - \mathbb{E}\,X^2\mathbf{1}_{|X|>\epsilon} \\
&= \mathbb{E}\,X^2 - 2\,\mathbb{E}\,X^2\mathbf{1}_{|X|>\epsilon}
\end{aligned}$$

and note $\operatorname{Var} X = \mathbb{E}\,X^2 \ge \mathbb{E}\,Y^2 \ge \operatorname{Var} Y$. Summing them over j, we have

$$\left|\mathbb{E}\,S_i\right| \le \sum_j \left|\mathbb{E}\,Y_{ij}\right| \le \frac{1}{\epsilon}L_i(\epsilon), 1 - 2L_i(\epsilon) \le \operatorname{Var} S_i \le 1$$

thus $\mathbb{E}\,S_n \to 0$ and $\operatorname{Var} S_n \to 1$ and claim follows from corollary 3.1.

$\square$

Chapter 4

Conditioning

4.1 Conditional Expectations

Throughout, X is an $\overline{\mathbb{R}}$-valued *r.v.* in $(\Omega, \mathcal{A}, \mathbb{P})$. Let $\mathcal{F}$ be sub-σ-algebra, and we regard it as a body of information. And the conditional expectation $\mathbb{E}_{\mathcal{F}} X = \mathbb{E}(X|\mathcal{F}) = \overline{X}$ is the "best" estimate of $X(\omega)$ given $\mathcal{F}$. Where the meaning of "best" comes from the usual MSE metric, $\mathbb{E}(X - \overline{X})^2$, which justified when X is square integrable.

4.1.1 Preparatory steps

Recall that $\mathcal{A}$ is the collection of all $\overline{\mathbb{R}}$-valued $r.v.'s$, similarly, $\mathcal{F}$ is the collection of all $\mathcal{F}$ measurable $r.v.'s$ taking value in $\overline{\mathbb{R}}$. WLOG, we assume $X \in \mathcal{A}_+$ for simplifying discussion.

Let A be an event containing ω, all we know about ω is it's in A. Based on this information, our best estimate should be simply the average over A:

$$\mathop{\mathbb{E}}_{H} X = \frac{1}{\mathbb{P}\, H} \int_H X d\,\mathbb{P} = \frac{\mathbb{E}\, X \mathbf{1}_H}{\mathbb{P}\, H}$$

where we assume $\mathbb{P}\, H > 0$. Such definition is agree with $\mathbb{E}\, X$ when $A = \Omega$. The number $\mathbb{E}_H X$ is called **conditional expectation** of X **given the event** A.

Next, suppose $\mathcal{F}$ is generated by a countable measurable partition $(A_i)_{i \in \mathbb{N}^*}$, that is, we know ω located in which one of them, thus

$$\mathop{\mathbb{E}}_{\mathcal{F}} X(\omega) = \sum_n (\mathop{\mathbb{E}}_{A_n} X) \mathbf{1}_{A_n}(\omega)$$

Where we regard $\mathbb{E}_{\mathcal{F}} X$ as a *r.v.* $\overline{X}$ and called **conditional expectation** of X given $\mathcal{F}$.

To proceed to general $\mathcal{F}$, note that:

1. $\overline{X} \in \mathcal{F}$.
2. $\mathbb{E}\, V X = \mathbb{E}\, V \overline{X}$ for every $V \in \mathcal{F}_+$. Which follows by noting V of the form $V = \sum a_n \mathbf{1}_{A_n}$.

4.1.2 Definition of conditional expectations

Definition 4.1. Let $\mathcal{F} \subset \mathcal{A}$, the **conditional expectation** of X given $\mathcal{F}$, denoted by $\mathbb{E}_{\mathcal{F}} H$ is:

1. If $X \in \mathcal{A}_+$. $\mathbb{E}_{\mathcal{F}} X$ is some *r.v.* $\overline{X}$ *s.t.*:

   1. $\overline{X} \in \mathcal{F}_+$
   2. $\mathbb{E}\, VX = \mathbb{E}\, V\overline{X}$

   And them we write $\mathbb{E}_{\mathcal{F}} X = \overline{X}$ and call $\overline{X}$ a version of $\mathbb{E}_{\mathcal{F}} X$.

2. For arbitrary X, if $\mathbb{E}\, X$ exists, define:

$$\mathop{\mathbb{E}}_{\mathcal{F}} X = \mathop{\mathbb{E}}_{\mathcal{F}} X^+ - \mathop{\mathbb{E}}_{\mathcal{F}} X^-$$

3. Otherwise, $\mathbb{E}_{\mathcal{F}} X$ is undefined, it's reasonable since so is $\mathbb{E}\, X$.

Remark.
1. The projection property is equivalent to $\mathbb{E}\, X\mathbf{1}_A = \mathbb{E}\, \overline{X}\mathbf{1}_A$ for any $A \in \mathcal{F}$.
2. $\mathbb{E}_{\mathcal{F}} X$ is unique up to equivalence.
3. Note $\mathbb{E}\, X = \mathbb{E}\, \mathbb{E}_{\mathcal{F}} X$ by letting $V = 1$, thus X is integrable iff so is $\mathbb{E}_{\mathcal{F}} X$. In which case, the projection property can be expressed as: for every $V \in \mathcal{F}_b$, $\mathbb{E}\, V(X - \overline{X}) = 0$.
4. Suppose X is integrable, then so is $\overline{X}$ and $\widetilde{X} = X - \overline{X}$. Then we have decomposition
$$X = \overline{X} + \widetilde{X}$$
where $\widetilde{X}$ is orthogonal to $\mathcal{F}$ *s.t.* $\mathbb{E}\, \widetilde{X}\mathbf{1}_A = 0$ for all $A \in \mathcal{F}$ and that's why we named "projection property".

4.1.3 Existence of conditional expectation

The following uses the Radon-Nikodym theorem to show the existence of conditional expectations. WLOG, we assume X is positive.

Theorem 4.1. *Let $X \in \mathcal{A}_+$ and $\mathcal{F} \subset \mathcal{A}$, then $\mathbb{E}_{\mathcal{F}} X$ exists and unique up to equivalence.*

Proof. $\forall A \in \mathcal{F}$, define

$$P(A) = \mathbb{P}\, A, Q(A) = \int_A X d\,\mathbb{P} = \mathbb{E}\, X\mathbf{1}_A$$

then P and Q are measures on $(\Omega, \mathcal{F})$ where $Q \ll P$ and we have $dQ = Xd\,\mathbb{P}$ by proposition 1.2. Hence, by the Radon-Nikodym theorem 1.19, there exists $\overline{X} \in \mathcal{F}_+$ *s.t.*

$$\mathbb{E}\, VX = \int_\Omega V dQ = \int_\Omega V\overline{X} dP = \mathbb{E}\, V\overline{X}$$

for every $V \in \mathcal{F}_+$. Thus the claim follows.

□

4.1.4 Examples

Example 4.1. Suppose our information is perfect, that is $X \in \mathcal{F}$, then $\mathbb{E}_{\mathcal{F}} X = X$ clearly.

Example 4.2. The other extreme case is $\mathcal{F}$ is independent with X. We claim that, $\mathbb{E}_{\mathcal{F}} X = \mathbb{E} X$. Check:

1. $\mathbb{E} X$ is a constant and thus $\mathcal{F}$-measurable clearly.
2. Note $\mathbf{1}_A$ and X are independent for $A \in \mathcal{F}$, then

$$\mathbb{E} X\mathbf{1}_A = \mathbb{E} X \, \mathbb{E} \mathbf{1}_A = \mathbb{E}(\mathbb{E} X\mathbf{1}_A)$$

Example 4.3. Suppose $(\Omega_i)_{i\in\mathbb{N}^*}$ is a at most countable partition of Ω where each has positive probability. Let $\mathcal{F}$ be σ-algebra generated by which. In the beginning of this chapter, we have

$$\underset{\mathcal{F}}{\mathbb{E}} X = \underset{\Omega_i}{\mathbb{E}} X = \frac{\mathbb{E} X\mathbf{1}_{\Omega_i}}{\mathbb{P}\,\Omega_i}$$

where Ω_i is where the outcome located.

Let $\mathbb{P}_{\mathcal{G}} A = \mathbb{P}(A|\mathcal{G}) = \mathbb{E}_{\mathcal{G}} \mathbf{1}_A$ and $\mathbb{P}_B A = \mathbb{P}(A|B) = \mathbb{E}_B \mathbf{1}_A$. Then

$$\mathbb{P} B \, \mathbb{P}(A|B) = \mathbb{P} B \underset{B}{\mathbb{E}} \mathbf{1}_A = \mathbb{E} \mathbf{1}_A \mathbf{1}_B = \mathbb{E} \mathbf{1}_{A\cap B} = \mathbb{P}(A \cap B)$$

and We have $\mathbb{P}(X|\mathcal{F}) = \mathbb{P}(X|\Omega_i)$.

Example 4.4. Suppose X, Y have joint distribution $f(x, y)$, in undergraduate probability, we have

$$\mathbb{E}(g(X)|Y = y) = \int g(x) \frac{f(x, y)}{\int f(x, y)dx} dx$$

Now we show that:

$$\underset{\sigma(Y)}{\mathbb{E}} g(X) = \mathbb{E}(g(X)|Y)$$

Clearly, $\mathbb{E}(g(X)|Y) \in \sigma(Y)$. For the projection properties, note if $A \in \sigma(Y)$ then $A = Y^{-1}(B)$ for some $B \in \mathcal{B}$, thus

$$\begin{aligned}
\mathbb{E}\left[(\mathbb{E}\, g(X)|Y)\mathbf{1}_A\right] &= \int_B \int g(x) \frac{f(x, y)}{f_Y(y)} dx f_Y(y) dy \\
&= \int_B \int g(x) f(x, y) dx dy \\
&= \mathbb{E}\, g(X)\mathbf{1}_{Y^{-1}(B)} = \mathbb{E}\, g(X)\mathbf{1}_A
\end{aligned}$$

4.1.5 Properties of conditional expectations

4.1.5.1 Similar to expectations

Proposition 4.1. *Assume all conditional expectations exists and $a, b, c \in \mathbb{R}$. Of course, all these conditional expectations exist if all the random variables are positive or integrable.*

- *Monotonicity:* $X \le Y \implies \mathbb{E}_{\mathcal{F}} X \le \mathbb{E}_{\mathcal{F}} Y$
- *Linearity:* $\mathbb{E}_{\mathcal{F}}(aX + bY + c) = a\,\mathbb{E}_{\mathcal{F}} X + b\,\mathbb{E}_{\mathcal{F}} Y + c$
- *MCT:* $X_n \ge 0, X_n \nearrow X \implies \mathbb{E}_{\mathcal{F}} X_n \nearrow \mathbb{E}_{\mathcal{F}} X$
- *Fatou's lemma:* $X_n \ge 0 \implies \mathbb{E}_{\mathcal{F}} \liminf X_n \le \liminf \mathbb{E}_{\mathcal{F}} X_n$
- *DCT: Suppose* $X_n \to X$ *and bounded by some integrable* Y*, then* $\mathbb{E}_{\mathcal{F}} X_n \to \mathbb{E}_{\mathcal{F}} X$
- *Jensen's inequality: Suppose* f *is convex, then* $\mathbb{E}_{\mathcal{F}} f(X) \ge f(\mathbb{E}_F X)$

Proof. **Monotonicity**. Note $X \le Y$ implies $\mathbb{E}\, X\mathbf{1}_A \le \mathbb{E}\, Y\mathbf{1}_A$ and that is $\mathbb{E}\, \overline{X}\mathbf{1}_A \le \mathbb{E}\, \overline{Y}\mathbf{1}_A$ and claim follows.

Linearity. Clearly the right side is $\mathcal{F}$ measurable and it's remain to show it's projection property. Then claim follows from linearity of integral and the projection property of $\mathbb{E}_{\mathcal{F}} X$ and $\mathbb{E}_{\mathcal{F}} Y$ by checking $V = \mathbf{1}_A$ for $A \in \mathcal{F}$.

MCT. Suppose $\overline{X}_n \nearrow \overline{X}$ where $\overline{X}_n$ is version of $\mathbb{E}_{\mathcal{F}} X_n$ respectively. Then it's remain to check $\overline{X}$ is version of $\mathbb{E}_{\mathcal{F}} X$. It's $\mathcal{F}$-measurable clearly and for $V \in \mathcal{F}_+$:

$$\mathbb{E}\, V\overline{X} = \lim_{n\to\infty} \mathbb{E}\, V\overline{X}_n = \lim_{n\to\infty} \mathbb{E}\, VX_n = \mathbb{E}\, VX$$

thus the claim follows.

□

4.1.5.2 Special properties

Proposition 4.2. *Let* $\mathcal{F}, \mathcal{G} \subset \mathcal{A}$ *and* $X, W \in \mathcal{A}$ *s.t.* $\mathbb{E}\, X$ *and* $\mathbb{E}\, WX$ *integrable, then*

1. *Conditional determinism:* $W \in \mathcal{F} \implies \mathbb{E}_{\mathcal{F}} WX = W\,\mathbb{E}_{\mathcal{F}} X$.
2. *Repeated conditioning:* $\mathcal{F} \subset \mathcal{G} \implies \mathbb{E}_{\mathcal{F}} \mathbb{E}_{\mathcal{G}} X = \mathbb{E}_{\mathcal{G}} \mathbb{E}_{\mathcal{F}} X = \mathbb{E}_{\mathcal{F}} X$.

Proof. **Conditional determinism**. WLOG, Suppose W is positive, then we have, for any $V \in \mathcal{F}_+$:

$$\mathbb{E}\, V(WX) = \mathbb{E}(VW)X = \mathbb{E}(VW)\overline{X} = \mathbb{E}\, V(W\overline{X})$$

as $VW \in \mathcal{F}_+$. It follows that $W\overline{X} = W\,\mathbb{E}_{\mathcal{F}} X$ is a version of $\mathbb{E}_{\mathcal{F}} WX$.

Repeated conditioning. $\mathbb{E}_{\mathcal{G}} \mathbb{E}_{\mathcal{F}} X = \mathbb{E}_{\mathcal{F}} X$ as example 4.1 by noting $\mathbb{E}_{\mathcal{F}} X \in \mathcal{F} \subset \mathcal{G}$. For the other side, note $\mathbb{E}_{\mathcal{F}} X \in \mathcal{F}$, for $A \in \mathcal{F} \subset G$:

$$\mathbb{E}\left[\mathbf{1}_A \mathop{\mathbb{E}}_{\mathcal{F}} X\right] = \mathbb{E}\,\mathbf{1}_A X = \mathbb{E}\left[\mathbf{1}_A \mathop{\mathbb{E}}_{\mathcal{G}} X\right]$$

by the projection properties of $\mathbb{E}_{\mathcal{G}} X$ and thus meet the projection properties of $\mathbb{E}_{\mathcal{F}}(\mathbb{E}_{\mathcal{G}} X)$.

□

Remark. For the repeated conditioning, think of $\mathcal{F}$ as the information a fool has, and $\mathcal{G}$ as that a genius has: the genius cannot improve on the fool's estimate, but the fool has no difficulty worsening the genius's. In repeated conditioning, fools win all the time.

Corollary 4.1. *If $\mathcal{F} \subset \mathcal{G}$ and $\mathbb{E}_{\mathcal{G}} X \in \mathcal{F}$ then $\mathbb{E}_{\mathcal{G}} X$ is a version of $\mathbb{E}_{\mathcal{F}} X$.*

Proof. $\mathbb{E}_{\mathcal{G}} X \in \mathcal{F}$ implies $\mathbb{E}_{\mathcal{F}} \mathbb{E}_{\mathcal{G}} X = \mathbb{E}_{\mathcal{G}} X$, which also equivalent to $\mathbb{E}_{\mathcal{F}} X$ by repeated conditioning.

□

4.1.6 Conditioning as projection

Recall we interpret $\mathbb{E}_{\mathcal{F}} X$ as a projection which minimize the MSE of X.

Theorem 4.2. *For every $X \in L^2(\mathcal{A})$, $\mathbb{E}_{\mathcal{F}} X$ minimize $\mathbb{E}\,|X - Y|^2$.*

Remark. $\overline{X}$ is the orthogonal projection of X onto $L^2(\mathcal{F})$ and we have decomposition $X = \overline{X} + \widetilde{X}$ where $\overline{X} \in L^2(\mathcal{F})$ and $\widetilde{X} \perp L^2(\mathcal{F})$.

Proof. Let $Z = \mathbb{E}_{\mathcal{F}} X - Y$ and $Y \in L^2(\mathcal{F})$ then:

$$\mathbb{E}(X - Y)^2 = \mathbb{E}(X - \mathop{\mathbb{E}}_{\mathcal{F}} X + Z)^2$$

It's sufficient to show that is $\mathbb{E}(X - \mathbb{E}_{\mathcal{F}} X)^2 + \mathbb{E}\, Z^2$. Which motivated us to consider cross-product term:

$$\begin{aligned}
\mathbb{E}\left[Z(X - \mathop{\mathbb{E}}_{\mathcal{F}} X)\right] &= \mathbb{E}\, ZX - \mathbb{E}(Z \mathop{\mathbb{E}}_{\mathcal{F}} X) \\
&= \mathbb{E}\, ZX - \mathbb{E}(\mathop{\mathbb{E}}_{\mathcal{F}} ZX) \\
&= \mathbb{E}\, ZX - \mathbb{E}\, ZX = 0
\end{aligned}$$

□

4.1.7 Conditional expectations given $r.v.'s$

Now we extend the definition in example 4.4. If $\{Y_t : t \in T\}$ is a collection of $r.v.'s$, then denoted $\mathcal{F} = \sigma\{Y_t : t \in T\}$ as usual. Then the conditional expectations given $\{Y_t : t \in T\}$ is just $\mathbb{E}_{\mathcal{F}} X$.

In the light of Doob-Dynkin theorem 2.13, we have $\mathbb{E}_{\sigma Y} X$ has the form $f \circ Y$.

4.2 Conditional probability and distribution

Recall definition in example 4.4:

$$\underset{\mathcal{F}}{\mathbb{P}} H = \underset{\mathcal{F}}{\mathbb{E}} \mathbf{1}_H = \begin{cases} \mathbf{1}_H & H \in \mathcal{F} \\ \mathbb{P} H & H \perp \mathcal{F} \end{cases}$$

4.2.1 Regular versions

Let $Q(A)$ be a version of $\mathbb{P}_{\mathcal{F}} A$ for each $A \in \mathcal{A}$. Clearly, $Q(\varnothing) = 0$ and $Q(\Omega) = 1$. Let $Q(\omega, A) = Q(A)(\omega) = Q_\omega(A)$, note

- $\omega \mapsto Q(\omega, A) = Q(A)$ is $\mathcal{F}$ measurable.
- $A \mapsto Q(\omega, A) = Q_\omega(A)$ is a measure:

  1. Nonnegativity: $Q_\omega(A) \geq 0$ for any A clearly and
  2. σ-additivity:

$$Q_\omega\left(\sum_n A_n\right) = \underset{\mathcal{F}}{\mathbb{E}} \mathbf{1}_{\sum_n A_n} = \underset{\mathcal{F}}{\mathbb{E}} \sum_n \mathbf{1}_{A_n} = \sum_n \underset{\mathcal{F}}{\mathbb{E}} \mathbf{1}_{A_n} = \sum_n Q_\omega(A_n)$$

However, the σ-additivity only enjoyed in a *a.s.* Ω_0 (as we use the MCT of conditional expectations which works on a *a.s.* set) and that keeps Q from being a transition kernel. Suppose Ω_a be the *a.s.* event *w.r.t.* sequence $a = (A_n)$. Then we need $\bigcap_a \Omega_a$ to be *a.s.* and this usually be a miserable object as there are uncountable many sequence a.

Nevertheless, it's often possible to pick versions of $Q(A)$ *s.t.* $\bigcap_a \Omega_a = \Omega$.

Definition 4.2. $Q(\omega, A)$ is said to be a **regular version** of $\mathbb{P}_{\mathcal{F}}$ or a **regular conditional probability** provided that Q is a transition probability kernel from $(\Omega, \mathcal{F})$ into $(\Omega, \mathcal{A})$.

The following is the reason for our interesting in regular version.

Proposition 4.3. *Suppose $\mathbb{P}_{\mathcal{F}}$ has a regular version Q, then*

$$QX(\omega) = Q_\omega X = \int X dQ_\omega$$

is a version of $\mathbb{E}_{\mathcal{F}} X$.

Proof. WLOG, assume $X \in \mathcal{A}_+$. By theorem 1.20, we have $QX \in \mathcal{F}_+$, then it's sufficient to check the projection property, to see this, suppose $X = \mathbf{1}_A$ for arbitrary $A \in \mathcal{A}$, then

$$\mathbb{E}\, VQX = \mathbb{E}\, VQ\mathbf{1}_A = \mathbb{E}\, VQ(A) = \mathbb{E}\, V\mathbf{1}_A = \mathbb{E}\, VX$$

then we can extends X to general case and claim follows.

□

The existence of regular version for $\mathbb{P}_{\mathcal{F}}$ require conditions either on $\mathcal{F}$ or $\mathcal{A}$.

- $\mathcal{F}$ generated by a measurable partition (Ω_n) of Ω, then

$$Q_\omega(A) = \sum_n \underset{\Omega_n}{\mathbb{P}} A \cdot \mathbf{1}_{\Omega_n}(\omega) = \underset{\Omega_i}{\mathbb{P}} A$$

  where Ω_i is where ω located and thus be a measure as so is $\mathbb{P}_{\Omega_i}$.