2025 Spring Math

HTOP,PDE,COMPLEX,ALGO

Yixia Yu

New York University

Edited by
Yixia Yu

Email: yy5091@nyu.edu

Contents

Chapter 1

Honors Theory of Probability

§1.1 Lecture 1 (02-03)Introduction to Probability

Office Hour: Wed 12:30-1:30pm,Fri 3:30-4:30pm W910

- Homework: weekly
- Grades: 5% participation, 15% homework, 40% midterm, final group project(Max 4 people a group,presetation 20%, Final report 20% 10 pages)

§1.1.1 Intro

Why do we need modern theory of probability?

Example 1

- Coin flip 7 times, what is P[first outcome and fifth outcome are Head]?
  sol:
  $$\Omega = (x1, x2, x3, x4, x5, x6, x7)$$
  $$A = (x1, x2, x3, x4, x5, x6, x7), x1 = x5 = H$$
  $$P(A) = \frac{|\Omega|}{|A|} = \frac{1}{4}$$
- Stock Price (mathematically,geometric Browmion motion, STochastic process, such that $t \to S_t$ is continuous but nowehere differentiable)
  What is P$[S_T \geq 100]$?
  $$\Omega = C[0, T] = \{\text{continuous function}, f[0, T] \to R\}$$
  $$A = \{f \in C[0, T] | f(t) \geq 100\}$$

  requires measure theory that defines P$[S_T \geq 100]$
  In modern Prob, Probability Space:(Ω, F, P), Ω is sample space, F is σ-algebra(meaningful subset of σ), P is probability measure($P : F \to [0, 1]$).

Topics covered:

- Probability space, σ-algebra, measure ,Conditional Probability and Independence
- Random variables (measurable functions),distribution
- Expectation (Lebesgue integral),Conditional distribution and expectation, functionos of random variables, Radom-Nikodym Derivative
- Random walks

- generating functions, characteristic functions
- Branching processes
- Convergence of random variables, Law of large numbers, Morte-Carlo Method
- Central Limit Theorem
- Time permitting: Large deviations, Markov Chains

§1.1.2 Probability Space

(Ω, F, P)

Example 2: Coinflip and Stock Price

- Coin flip infinite times,

$$\Omega = \{(x1, x2, \cdots), x_i = H, T\}$$

Let A be the event of 10^6 consecutive Tails,

$$A = \bigcup_{i=1}^{+\infty} \{x_i = x_{i+1} = \cdots = x_{i+10^6-1} = 0, (x_i \in \Omega)\}$$

$P[A] = 1$

- Discrete Stock Price model, t=0,1,2,$\cdots$,T
T is maturing time, time step $\Delta t \ll T$

$$N = \frac{T}{\Delta t}$$

$$\text{price go} \begin{cases} \nearrow \text{by factor} : e^{\sigma\sqrt{\Delta t}} \\ \searrow \text{by factor} : e^{-\sigma\sqrt{\Delta t}} \end{cases}$$

$$\Omega = \{(x_1, x_2, \cdots, x_N), x_i = 0 \text{ or } 1\}$$

Stock price at time t: $\forall \omega \in \Omega$

$$S_N(\omega) = S_0 e^{\sum_{i=1}^N x_i \sigma\sqrt{\Delta t} e^{(N - \sum_{i=1}^N x_i)(-\sigma\sqrt{\Delta t})}}$$

$$S : \Omega \to R(\text{Random Variable})$$

Event: return at T is positive but not more than 10%:

$$\{\omega \in \Omega : S_N(\omega) \in (S_0, 1.1S_0]\}$$

Example 3: Gambling

- Gambling:

- start with {0,1,2,$\cdots$} each time bet an interger amount
- if amount of money =0, stays at 0
  wealth process: $\Omega - \{0,1,2,\cdots\} \times \{0,1,2,\cdots\} \times \cdots$
  wealth after time n: (random variable) $X_n : \Omega \to N, (x_1, x_2, ...x_n) \to x_n$
  $(X_n)$is a Markov Chain (future only depends on present state, but not the past)

$$\begin{aligned}&\text{Event: \{State j is reached from state i\}}\\ &= \{\omega \in \Omega : \exists n \in N, X_0(\omega) = i, X_n(\omega) = j\}\\ &= \bigcup_{m=1}^{+\infty} \{\omega \in \Omega : X_o(\omega) = i, X_n(\omega) = j\}\end{aligned}$$

§1.2 Lecture 2 (02-05) – Algebra

§1.2.1 algebra and σ-algebra

in practice, want F to be closed under $\bigcap, \bigcup, ^c$

Definition 1.2.1

Let $\mathcal{A}$ be a collection of subsets of Ω , $\mathcal{A}$ is an algebra iff:

- $\Omega \in \mathcal{A}$
- if A,B$\in \mathcal{A}$ then A$\cup$B $\in \mathcal{A}$
- if A$\in \mathcal{A}$ then A$^c \in \mathcal{A}$

Remark 1

if A,B$\in \mathcal{A} \to$ A $\cap$ B$\in \mathcal{A}$, because A$\cap$B = (A$^c\cup$B$^c)^c$

Fact:

① $P(\Omega)$(powerset)={A:A $\subset \Omega$} is an algebra

② smallest algebra/trivial algebra: $\{\emptyset, \Omega\}$

③ Let $\mathcal{A}_1, \mathcal{A}_2$ be two algebras of Ω

$$\mathcal{A}_1 \cap \mathcal{A}_2 = \{B \in \Omega : B \in \mathcal{A}_1 \text{ and } B \in \mathcal{A}_2\} \text{ is an algebra}$$

$$\text{if}(\mathcal{A}_j)_{j\in J}\text{is a family of algebras, then } \bigcap_{j\in J} \mathcal{A}_j \text{ is an algebra}$$

④ Let $\mathcal{E}$ be any collection of subsets of Ω

$$a(\mathcal{E}) = \bigcap_{\substack{\mathcal{A}\supseteq\mathcal{E}\\ \mathcal{A}\text{ is an algebra}}} \mathcal{A} \text{ is an algebra}$$

$a(\mathcal{E})$ is an algebra by ③
in fact, $a(\mathcal{E})$ is the smallest algebra containing $\mathcal{E}$
It is called the algebra generated by $\mathcal{E}$

⑤ Let $A \subseteq \Omega$,$\mathcal{E} = \{A\}$
Then

$$a(\mathcal{E}) = \{\underbrace{A, A^c, \Omega, \emptyset}_{f}\}$$

Proof:

- $a(\mathcal{E}) \subseteq f$
  notice that f is an algebra since $f \supseteq \mathcal{E}$, therefore $f \supseteq a(\mathcal{E})$ because $a(\mathcal{E})$ is the smallest algebra containing $\mathcal{E}$.
- $f \subseteq a(\mathcal{E})$:

$$A \in a(\mathcal{E}), A^c \in a(\mathcal{E})$$

  beacuse $a(\mathcal{E})$ is an algebra, $\emptyset, \Omega \in a(\mathcal{E})$

⑥ $\pi = \{A_1, A_2, \cdots, A_n\}$, $\Omega = \bigcup_{i=1}^n A_i, A_i \cap A_j = \emptyset$ Then:

$$a(\pi) = \{\bigcup_{i \in I} A_i, for I \subset 1, 2, \cdots, n\} = \text{ finite disjoint union of } (A_i)_{i=1}^n$$

⑦ Let $\mathcal{A}$ be an algebra of $\mathbb{R}$
Let $X : \Omega \to$ be a function
Then

$$\underbrace{\{X^{-1}(A), A \in \mathcal{A}\}}_{\{\omega \in \Omega : X(\omega) \in A, \text{ for some } A \in \mathcal{A}\}} \qquad \text{is an algebra of } \Omega$$

Hint:

$$X^{-1}(A \cup B) = (X^{-1}(A)) \cup (X^{-1}(B))$$

⑧ $\Omega = \mathbb{R}$, $\mathcal{E} = \{\text{left open right closed intervals}\} = \begin{cases} (a, b], & -\infty \leq a < b < +\infty \\ (a, +\infty) \end{cases}$

Then:

$$\begin{aligned} a(\mathcal{E}) &= \text{"finite disjioint union of elements in } \mathcal{E}\text{"} \\ &= \underbrace{\{I_1 \cup \cdots \cup I_k; I_j \in \mathcal{E}, I_i \cap I_j = \emptyset\}}_{f} \end{aligned}$$

Hint:

- $f \subseteq a(\mathcal{E})$ strightforward
- $a(\mathcal{E}) \subseteq f$
  - check f is an algebra
  - since $\mathcal{E} \subset f$, $a(\mathcal{E})$ is the smallest algebra containing $\mathcal{E}$, $a(\mathcal{E}) \subseteq f$
  - $(a, b]^c = \underbrace{(-\infty, a] \cup (b, +\infty)}_{\in f}$

Lemma 1.2.1

In Probability, if $(A_k)_{k=1}^n$ are disjoint events, we have $P(\bigcup_{k=1}^{+\infty} A_k) = \sum_{k=1}^{+\infty} P(A_k)$

Definition 1.2.2: $\sigma - algebra$

A $\sigma - algebra$ $(\sigma - field)$ $\mathcal{A}$ is an collection of subsets of Ω such that

- $\Omega \in \mathcal{A}$
- if $A_1, A_2, \cdots \in \mathcal{A}$ then $\bigcup_{i=1}^{+\infty} A_i \in \mathcal{A}$
- if $A \in \mathcal{A}$, then $A^c \in \mathcal{A}$

Remark 2

if $A_1, A_2, \cdots \in \mathcal{A}$ then $\bigcap_{i=1}^{+\infty} A_i \in \mathcal{A}$
σ -algebra represents the collection of information determined by partial derivatives

Example 1

- coinflips infinity many times

$$\Omega = \{(x_1, x_2, \cdots)x_i = 0, 1\} = (0, 1)^\infty$$

After observing first outcome, $f_1 = \{\emptyset, \Omega, A_0, A_1\}$
Where $A_0 = \{(0, x_2, x_3, \cdots), x_i = 0, 1\}, A_1 = \{(1, x_2, x_3, \cdots), x_i = 0, 1\}$

After observing second outcome, $f_2 = \{\emptyset, \Omega, A_0, A_1, A_{00}, A_{01}, A_{10}, A_{11}\}$
Where $A_{00} = \{(0, 0, x_3, x_4, \cdots), x_i = 0, 1\}, A_{01} = \{(0, 1, x_3, x_4, \cdots), x_i = 0, 1\} \cdots$

After observing first n outcomes,

$$f_n = \{\emptyset, \Omega, (A_j)_{j \in (\sigma 1)^n} \text{and finite disjioint unions}\}$$

and

$$f_1 \subseteq f_2 \subseteq \cdots \subseteq f_n \subseteq \cdots$$

Proposition 1.2.0

1. intersection of σ-algebras is a σ-algebra
2. Let $\mathcal{E}$ be a collecion of subsets of Ω

$$\sigma(\mathcal{E}) = \bigcap_{\substack{f \supseteq \mathcal{E} \\ f \text{ is an algebra}}} f \text{ is a smallest } \sigma\text{-algebra that contains } \mathcal{E}$$

3. For any collection $\mathcal{E}$, we have $a(\mathcal{E}) \subseteq \sigma(\mathcal{E})$

④ For any collection $\mathcal{E}$, we have $\sigma(a(\mathcal{E})) = \sigma(\mathcal{E})$
Hint:

$$a(\mathcal{E}) \subseteq \sigma(\mathcal{E}) \to \sigma(a(\mathcal{E})) \subseteq \sigma(\mathcal{E})$$
$$\mathcal{E} \subseteq \sigma(a(\mathcal{E})) \to \sigma(\mathcal{E}) \subseteq \sigma(a(\mathcal{E}))$$

§1.3 Recitation 1 (02-07) – Problem Solving

§1.3.1 Basic Set Theory

Definition 1.3.1: $\cup, \cap, ^c$

De Morgan's Law:

$$(\bigcup_{j\in J} A_j)^c = \bigcap_{j\in J} A_j^c$$
$$(\bigcap_{j\in J} A_j)^c = \bigcup_{j\in J} A_j^c$$

Definition 1.3.2: $\backslash$

$A\backslash B = A \cap B^c$
Then $A\backslash B = A\cap B^c = A\backslash(A\cap B) = B^c\backslash A^c$

Remark 1

$$\bigcap_{n\geq 1} A_n = A_1\backslash(\bigcup_{j\geq 2}(A_1 - A_j)) \text{ (ex)}$$

§1.3.2 Limits of Sets

Definition 1.3.3: Limit Sets

Let $(A_n)_{n\geq 1}$ be a sequence of sets, then

$$B_k = \bigcup_{n\geq k} A_n, C_k = \bigcap_{n\geq k} A_n$$

Then B_k is increasing, C_k is decreasing
Define:

$$\limsup_{n\to+\infty} A_n = \lim_{k\to+\infty} B_k = \bigcap_{k\geq 1}\bigcup_{n\geq k} A_n$$
$$\liminf_{n\to+\infty} A_n = \lim_{k\to+\infty} C_k = \bigcup_{k\geq 1}\bigcap_{n\geq k} A_n$$

Definition 1.3.4: liminf, limsup of sequence

$$\limsup an = \inf_{k\geq 1}\sup_{n\geq k} a_n, \liminf an = \sup_{k\geq 1}\inf_{n\geq k} a_n$$

When $\limsup A_n = \liminf A_n$, Then we say $\lim_{n\to+\infty} A_n$ exist and $\lim_{n\to+\infty} A_n = \limsup A_n = \liminf A_n$

In probability,

$$\begin{aligned}
\limsup A_n &= \{A_n \text{ occurs infinitely often}\} \\
&= \{An.i.o\} \\
x \in \limsup A_n &\Leftrightarrow \forall k \in \mathbb{N}, \exists n \geq k \text{ such that } x \in A_n \\
&\Leftrightarrow A_n.i.o \\
\liminf A_n &= \{A_n \text{ occurs eventually}\} \\
&\Leftrightarrow \exists k \in \mathbb{N}, \forall n \geq k, x \in A_n
\end{aligned}$$

Remark 2

1. if A_n increases, then $\limsup A_n = \lim_{n\to\infty} A_n = \bigcup_{n\geq 1} A_n$
   if A_n decreases, then $\liminf A_n = \lim_{n\to\infty} A_n = \bigcap_{n\geq 1} A_n$
2. $$(\limsup A_n)^c = \liminf A_n, (\liminf A_n)^c = \limsup A_n$$

§1.3.3 Exercise

1. $A_k = \begin{cases} E, \text{if k is odd} \\ F, \text{if k is even} \end{cases}$
   Then $\limsup A_n = E \cup F, \liminf A_n = E \cap F$
2. Let $f_n : \mathbb{R} \to \mathbb{R}$, Let A=$\{x \in \mathbb{R} : \lim_{n\to\infty} f_n(x) = f(x)\}$ Then,
   $$\begin{aligned}
   A^c &= \bigcup_{k=1}^{+\infty}\bigcap_{m=1}^{+\infty}\bigcup_{n=m}^{+\infty}\{x : |f_n(x) - f(x)| \geq \frac{1}{k}\} \\
   &= \bigcup_{k=1}^{+\infty}[\limsup_n\{x : |f_n(x) - f(x)| \geq \frac{1}{k}\}]
   \end{aligned}$$
3. Suppose that $\lim_{n\to\infty} f_n(x) = f(x), \forall x \in \mathbb{R}$
   Then
   $$\begin{aligned}
   \{x : f(x) \leq t\} &= \bigcap_{k=1}^{+\infty}\bigcup_{m=1}^{+\infty}\bigcup_{n=m}^{+\infty}\{x \in \mathbb{R} : f_n(x) < t + \frac{1}{k}\} \\
   &= \bigcap_{k=1}^{+\infty}[\liminf_n\{x : f_n(x) < t + \frac{1}{k}\}]
   \end{aligned}$$

Proof:

1. ex2: Want

$$A = \bigcup_{k\geq 1}^{+\infty} \bigcap_{\geq =1}^{+\infty} \bigcup_{n\geq m}^{+\infty} \{x : |f_n(x) - f(x)| \geq \frac{1}{k}\}$$

$$f_n(x) \to f(x) \text{ iff } \forall\epsilon > 0, \exists N \geq N, |f_n(x) - f(x)| < \epsilon$$

$$\{x : f_n(x) \to f(x)\} = \bigcap_{\epsilon>0} \bigcup_{N} \bigcap_{n\geq N} \{x : |f_n(x) - f(x)| < \epsilon\},$$

$$\text{USE that } \{|f_n - f| < \epsilon\} \text{ is monotone increasing in } \epsilon$$

Review on mapping:

$$f : X \to Y, f^{-1} : Y \to X$$

Basic properties:

- $f(\bigcup_{i\in I}) = \bigcup_{i\in I} f(A_i)$
- $f(\bigcap_{i\in I}) = \bigcap_{i\in I} f(A_i)$

For (B_i) subset of Y:

- if $B_1 \subset B_2, f^{-1}(B_1) \subset f_{-}1(B_2)$
- $f^{-1}(\bigcup_{i\in I} B_i) = \bigcup_{i\in I} f^{-1}(B_i)$
- $f^{-1}(B^c) = (f^{-1}(B))$

Definition 1.3.5: Indicator Mapping

$$1_A : X \to \{0, 1\}$$

$$1_A(x) = \begin{cases} 1, x \in A \\ 0, x \notin A \end{cases}$$

Exercise:

1. $1_{limsupA_n} = limsup 1_{A_n}$
2. $1_{liminfA_n} = liminf 1_{A_n}$

Proof:
if

$$\begin{aligned} 1_{limsupA_n}(x) = 1 &\Leftrightarrow x \in limsupA_n \\ &\Leftrightarrow \forall k \in \mathbb{N}, \exists n \geq k, x \in A_n \\ &\Leftrightarrow \forall k \in \mathbb{N}, \exists n \geq k, 1_{A_n}(x) = 1 \\ &\Leftrightarrow limsup 1_{A_n}(x) \geq 1 (\text{Definition of limsup}) \\ &\Leftrightarrow 1_{limsupA_n}(x) = 1 \end{aligned}$$

Exercise: Let $\mathcal{A}_1, \mathcal{A}_2$ be algebras of Ω

1. show that $\underbrace{\mathcal{A}_1 \cap \mathcal{A}_2}_{B\subset\Omega : B\in\mathcal{A}_1, B\in\mathcal{A}_2}$ is an algebra

2. show that $\underbrace{\mathcal{A}_1 \cup \mathcal{A}_2}_{B\subset\Omega: B\in\mathcal{A}_1 \text{ or } B\in\mathcal{A}_2}$ is an algebra iff $\mathcal{A}_1 \subseteq \mathcal{A}_2$ or $\mathcal{A}_2 \subseteq \mathcal{A}_1$

Proof: Suppose by contradiction that

$$\exists A_1 \in \mathcal{A}_1 \text{ but } A_1 \notin \mathcal{A}_2 \text{ and } A_2 \in \mathcal{A}_2 \text{ but } A_2 \notin \mathcal{A}_1 \text{ and } \mathcal{A}_1 \cup \mathcal{A}_2 \text{ is an algebra}$$

Therefore:

$$A_1 \cup A_2 \in \mathcal{A}_1 \cup \mathcal{A}_2, A_1 \backslash A_2 = A_1 \cap \underbrace{A_2^c}_{\mathcal{A}_1\cup\mathcal{A}_2} \in \mathcal{A}_1 \cup \mathcal{A}_2$$

$$A_2 \backslash A_1 = A_2 \cap A_1^c \in \mathcal{A}_1 \cup \mathcal{A}_2$$

$\Rightarrow$ at least two of

$(A_1\backslash A_2)\cup(A_2\backslash A_1), A_1\backslash A_2, A_2\backslash A_1$ are in $\mathcal{A}_1$ or $\mathcal{A}_2$

Assume in $\mathcal{A}_1$

$\Rightarrow$ (by $\mathcal{A}_1$ is an algebra, all three sets are in $\mathcal{A}_1$)

$$\Rightarrow A_2 = (\underbrace{A_1 \cup A_2}_{\in\mathcal{A}_1})\backslash(\underbrace{A_1\backslash A_2}_{\mathcal{A}_1}) \in \mathcal{A}_1$$

Contradiction!

§1.4 Lecture 3 (02-10)–(Content and Measure)

Recall σ-algebra:

Example 1

① We know $\Omega = \mathbb{R}$,

$$\mathcal{E} = \{\text{left open right closed intervals}\} = \begin{cases} (a,b], & -\infty \le a < b < +\infty \\ (a,+\infty), & a \in \mathbb{R} \end{cases}$$

Then we know:

$$a(\mathcal{E}) = \{ \text{"finite disjioint union of elements in } \mathcal{E}\text{"}\}$$

What is $\sigma(a(\mathcal{E}))$?
$\sigma(\epsilon) =$ Borel Sets $\mathcal{B}(R)$
Any "reasonable" subset of $\mathbb{R}$ is in $\sigma(\epsilon)$

- $(a,b) \in \sigma(\epsilon) : (a,b) = \bigcup_{n\ge1} \underbrace{(a, b-\frac{1}{n}]}_{\in\sigma(\epsilon)} \in \sigma(\epsilon)$
- any singleton $\{a\} \in \sigma(\epsilon)$, because $\{a\} = \bigcap_{n\ge1} \underbrace{(a-\frac{1}{n}, a+\frac{1}{n})}_{\in\sigma(\epsilon)} \in \sigma(\epsilon)$
- any countable set is in $\sigma(\epsilon)$ (e.g. $\mathbb{Q}$)
- The set of trancsendental numbers is in $\sigma(\epsilon)$, because the set of algebraic numbers is countable

Definition 1.4.1: measurable set

A pair (Ω, F) , where F is a σ-algebra of Ω, is called a measurable space
Any set $A \in F$ is called a measurable set

§1.4.1 Content and Measure

Definition 1.4.2: Content

Let $\mathcal{A}$ be an algebra of Ω, A set function $\mu : \mathcal{A} \to [0, +\infty)$ is called a content iff:

- $\mu(\emptyset) = 0$
- if A,B$\in \mathcal{A}$ and $A \cap B = \emptyset$, then $\mu(A \cup B) = \mu(A) + \mu(B)$ (finite additivity)

Lemma 1.4.1

Let $\mu : \mathcal{A} \to [0, +\infty)$ be a content, $\forall A, B \in \mathcal{A}$ then:

① $\mu(A) + \mu(B) = \mu(A \cup B) + \mu(A \cap B)$

② if $A \subset B$, and $\mu(A) < +\infty$,then $\mu(B \backslash A) = \mu(B) - \mu(A)$

③ if $A \subset B$, then $\mu(A) \leq \mu(B)$

④ if $A_1, A_2, \cdots \in \mathcal{A}$, then $\mu(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} \mu(A_i)$

⑤ if $A_1, A_2, \cdots \in \mathcal{A}$, $A_i \cap A_j = \emptyset$ and $\bigcup_{i=1}^{+\infty} \in A$ then $\mu(\bigcup_{i=1}^{+\infty} A_i) \geq \sum_{i=1}^{+\infty} \mu(A_i)$

Proof:
finite add:
1:

$$\begin{aligned} \mu(B) &= \mu(A \cup B) + \mu(B \backslash A) \\ \mu(A) + \mu(B \backslash A) &= \mu(A \cup B) \\ \Rightarrow \mu(A) + \mu(B) &= \mu(A \cup B) + \mu(A \cap B) \end{aligned}$$

4:

$$\text{Let } B_1 = A_1, B_2 = A_2 \backslash A_1, B_3 = A_3 \backslash (A_1 \cup A_2), \cdots$$
$$\text{Then } B_j \text{ are disjoint and } \bigcup_{i=1}^{n} B_i = \bigcup_{i=1}^{n} A_i$$
$$\text{By finite add for } B_j :$$
$$\mu(\bigcup_{i=1}^{n} A_i) = \mu(\bigcup_{i=1}^{n} B_i) = \sum_{i=1}^{n} (\mu(A_i) \backslash \bigcup_{i=1}^{j=1} A_j) \leq \sum_{i=1}^{n} \mu(A_i)$$

5:

For $\forall n$,

$$\mu(\bigcup_{i=1}^{+\infty} A_i) \geq \mu(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} \mu(A_i) \text{ send n } \nearrow +\infty \text{ to conclude}$$

Remark 1

In general. $\mu(\bigcup_{i=1}^{+\infty} A_i) \neq \sum_{i=1}^{+\infty} \mu(A_i)$ although $\mu(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} \mu(A_i)$ require continuity of μ

Counterexample:

- Let $\Omega = \mathbb{R}, \mathcal{A} = a(\epsilon)$, $\forall A \in \mathcal{A}$,

  $$\mu(A) = \lim_{L \to +\infty} \frac{\overbrace{|A \cup [0, L]|}^{\text{Length of interval}}}{L} \text{ "Density of A in } [0, +\infty)\text{"}$$

  Then μ is a content.
  Take $A_i = (i, i+1]$, then $\mu((i, i+1]) = 0$, But $\mu(\bigcup_{i=0}^{+\infty}) = \mu((0, +\infty)) = 1$

- For $\mathbb{R}, a(\epsilon)$, Given $A \in a(\epsilon)$

  $$\mathring{A} = \{x \in A : \exists r_x > 0, s.t. (x - rx, x + rx) \subset A\}$$
  $$\partial A = \bar{A} \backslash \mathring{A}$$

  $$\mu(A) = \begin{cases} 2, & \text{if } 0 \in \mathring{A} \\ 1, & \text{if } 0 \in \partial A \\ 0, & \text{else} \end{cases}$$

  Then μ is a content
  However, $A_i = (\frac{1}{i+1}, \frac{1}{i}], \mu(A_i) = 0$, But $\mu(\bigcup_{i=1}^{+\infty}) = \mu((0, 1])$

Example:

- (Discrete Probability)

  $$\Omega = \{\omega_1, \cdots, \omega_n\} \text{ finite set, } , F = P(\Omega)$$

  Set $A_i = \{\omega_i\}, P(A_i) = P(\omega_i) = P_i$ such that $\sum_{i=1}^{n} P_i = 1$
  Then $P : P(\Omega) \to [0, 1]$ defines a content on $P(\Omega)$ by extending the P using finite additivity:

  $$\forall A \in P(\Omega), P(A) = \sum_{\omega \in A} P(\omega)$$

- $\Omega = \mathbb{R}$, algebra $\mathcal{A} = a(\epsilon) = \{$ finite disjoint umiom of elements in $\epsilon\}$
  Define $m : a(\epsilon) \to [0, +\infty)$, set $m([a, b])$ =b-a,
  and extend by additivity:

  $$m(I) = \sum_{i=1}^{n} m(I_j), \text{ if } I = I_1 \cup \cdots \cup I_n, I_j \cap I_i = \emptyset$$

$\Rightarrow$ m is a content on $(\mathbb{R}, a(\epsilon))$
m can be further extended to $(\mathbb{R}, \sigma(\epsilon))$, called Lebesgue Measure

Definition 1.4.3: countably additive

A content $\mu : \mathcal{A} \to [0, +\infty)$ is countably additive if:

$$\mu(\bigcup_{i=1}^{+\infty} A_i) = \sum_{i=1}^{+\infty} \mu(A_i) \text{ for every disjoint } A_1, A_2, \dots \in \mathcal{A}$$

Definition 1.4.4: measure

Let (Ω, F) be an measurable space, then a content $\mu : F \to [0, +\infty)$ that is countably additive is called a measure

Lemma 1.4.2

$m : a(\epsilon) \to [0, +\infty)$ is countably additive

Proof:
Let $A_1, A_2, \dots \in a(\epsilon), (A_k)$ disjioint , $A := \bigcup_{i=1}^{+\infty} A_i \in a(\epsilon)$
Want to show :

$$m(A) = \sum_{i=1}^{+\infty} m(A_i)$$

we can write, using $A, (A_k) \in a(\epsilon)$, $A = \bigcup_{j=1}^{n} I_j$, where $I_j \in \epsilon$ and (I_j) disjioint
$A_j = \bigcup_{k=1}^{n_i} J_{ik}$, where $J_{ik} \in \epsilon$,(J_{ik}) disjioint
Then:

$$\begin{aligned}
m(A) &= \sum_{j=1}^{n} m(I_j) \\
&= \sum_{j=1}^{n} \sum_{j=1}^{+\infty} \sum_{k=1}^{n_i} m(I_j \cap J_{ik}) \\
&= \sum_{i=1}^{+\infty} m(\underbrace{\bigcup_{j=1}^{n} I_j}_{A} \cap (\underbrace{\bigcup_{k=1}^{n_i} J_{i_{jk}}}_{A_i})) \\
&= \sum_{i=1}^{+\infty} m(A_i)
\end{aligned}$$

§1.5 Lecture 4 (02-12)–Measure and Extension

$$\Omega = \mathbb{R}, m : \epsilon \to [0, +\infty), \text{ such that } \begin{cases} m([a,b]) = b - a, \\ m((a, +\infty)) = +\infty \end{cases}$$

$$\text{extend m to } a(\epsilon) : \forall A \in a(\epsilon),$$

$$\text{if } A = \bigcup_{i=1}^{n} I_j, I_j \text{ disjioint}, \ m(A) = \sum_{j=1}^{n} m(I_j)$$

Fact:
if $I \in \epsilon$ s.t. $I = \bigcup_{i=1}^{+\infty} I_i, (I_i)$ disjoint and $I_j \in \epsilon$
Then

$$\begin{aligned} m(I) &= |I| \\ &= \sum_{i=1}^{+\infty} |I_i| = \bigcup_{i=1}^{+\infty} m(I_i) \end{aligned}$$

Lemma 1.5.1

we want to prove m is a countably additive content on $(\mathbb{R}, a(\epsilon))$

$$\text{Let } A_j \in a(\epsilon), A = \bigcup_{j=1}^{\infty} \in a(\epsilon), (A_j) \text{ disjoint}$$

$$\exists (I_i) \text{ such that } I_i \in \epsilon, \ \text{disjoint } A = \bigcup_{i=1}^{n} I_i$$

$$\exists (J_{ij}) \text{ such that } J_{ij} \in \epsilon, \ \text{disjoint } A_j = \bigcup_{k=1}^{j} J_{ik}$$

$$\begin{aligned} m(A) \underbrace{=}_{\text{def}} \sum_{i=1}^{n} m(I_i) &= \sum_{i=1}^{n} m(\bigcup_{j,k} \underbrace{(I_i) \cap (J_{ik})}_{\in \epsilon}) \\ &\underbrace{=}_{\text{fact}} \sum_{i=1}^{n} \sum_{j=1}^{+\infty} \sum_{k=1}^{n_j} m(I_i \cap I_{jk}) \\ &\underbrace{=}_{\text{finite add}} \sum_{j=1}^{+\infty} m(A_j) \end{aligned}$$

Theorem 1.5.1

m extends to a measure on $(\mathbb{R}, \sigma(\epsilon) = \mathcal{B}(\mathbb{R}))$ It is the unique measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ such that

$$m([a, b]) = b - a$$

Definition 1.5.1

If μ is a measure on (Ω, F), then (Ω, F, μ) is said to be a measure space
If $\mu(\Omega) = 1$, then (Ω, F, μ) is said to be a probability space

Lemma 1.5.2

Let (Ω, F, μ) be a measure space, then:

① stability: Let $A_1, A_2, \cdots \in F$, then $\mu(\bigcup_{i=1}^{+\infty} A_i) \leq \sum_{i=1}^{+\infty} \mu(A_i)$

② continuity from below: Let $A_1, A_2, \cdots \in F$, $A_1 \subseteq A_2 \subseteq \cdots$, then

$$\mu(\bigcup_{i=1}^{+\infty} A_i) = \lim_{i \to +\infty} \mu(A_i) = \mu(\lim_{i \to +\infty} A_i)$$

③ continuity from above: Let $A_1, A_2, \cdots \in F$, $A_1 \supseteq A_2 \supseteq \cdots$,
$(\mu(A_i) < +\infty$ we need this for the Counterexample $A_i = [i, +\infty)$) then

$$\mu(\bigcap_{i=1}^{+\infty} A_i) = \lim_{i \to +\infty} \mu(A_i) = \mu(\lim_{i \to +\infty} A_i)$$

Proof:

① Let $B_1 = A_1, B_2 = A_2 \backslash A_1, \cdots, B_n = A_n \backslash \bigcup_{i=1}^{n-1} A_i$,
Then (B_i) disjoint, $\bigcup_{i=1}^{+\infty} A_i = \bigcup_{i=1}^{+\infty} B_i$

$$\mu(\bigcup_{i=1}^{+\infty} A_i) = \mu(\bigcup_{i=1}^{+\infty} B_i) = \sum_{i=1}^{+\infty} \mu(B_i) = \sum_{i=1}^{+\infty} \mu(A_i \backslash \bigcup_{j=1} i - 1 A_j) \leq \sum_{i=1}^{+\infty} \mu(A_i)$$

② Let $B_1 = A_1, B_2 = A_2 \backslash A_1, \cdots, B_n = A_n \backslash A_{n-1} \cdots$,

$$\begin{aligned} \mu(\bigcup_{i=1}^{+\infty} A_i) = \mu(\bigcup_{i=1}^{+\infty} B_i) = \sum_{i=1}^{+\infty} \mu(B_i) \\ &= \mu(A_1) + \sum_{i=2}^{+\infty} (\mu(A_i) - \mu A_{i-1}) \\ &= \lim_{n \to +\infty} (\mu(A_1) + \sum_{i=2}^{n} (\mu(A_i) - \mu A_{i-1})) \\ &= \lim_{n \to +\infty} \mu(A_n) \end{aligned}$$

③

$$\mu(A_1) - \mu(\bigcap_{i=1}^{+\infty} A_i) = \mu(A_1 \backslash \bigcap_{i=1}^{+\infty} A_i) = \mu(\bigcup_{i=1}^{+\infty} A_1 \backslash A_i)$$

Since $A_1 \nearrow$, by $2 := \lim_{n \to +\infty} (\mu(A_1) - \mu(A_i)) = \mu(A_1) - \lim_{n \to +\infty} \mu(A_i)$

Definition 1.5.2: σ-finite measure

Given a measure space (Ω, F, μ), μ is said to be finite if $\mu(\Omega) < +\infty$
μ is σ-finite if there exists $(E_i)_{i=1}^{+\infty}$ such that $\bigcup_{i=1}^{+\infty} E_i \in \Omega$ and $\mu(E_i) < +\infty$

Example 1

$(\mathbb{R}, \mathcal{B}(\mathbb{R}), m)$ is σ-finite

Definition 1.5.3

if $F \in \mathcal{F}$ is such that $\mu(F) = 0$, then F is called a μ-null set

Example 2

$m(\{a\}) = 0$ because

$$m(\{a\}) = m(\bigcap_{n \geq 1} (a - \frac{1}{n}, a]) = \text{(continuous from above)} \lim_{n \to +\infty} m((a - \frac{1}{n}, a]) = 0$$

$$m(Q) = \sum_{q \in Q} m(\{q\}) = 0$$

Recall:
Start with $(R, a(\epsilon), m)$ m is content + countably additive
extention: $(R, a(\epsilon), m) \to (R, \sigma(\epsilon), m)$ m is measure

Theorem 1.5.2: Caratheodory Extention Theorem

Let F be an algebra on Ω , μ be a countably additive content on (Ω, F),
If μ is σ-finite, then μ extends to a measure on $(\Omega, \sigma(F))$

Example 3

Let $\epsilon = \begin{cases} (a, b], & -\infty \leq a < b < +\infty \\ (a, +\infty), & a \in \mathbb{R} \end{cases}$ Let $m_F : a(\epsilon) \to [0, +\infty)$ be a content, such that

$$m_F([a, b]) = F(b) - F(a), m_F((a, +\infty)) = F(+\infty) - F(a)$$

Where F is a right continuous increasing function on $\mathbb{R}$,

$$F(\pm\infty) = \lim_{x \to \pm+\infty} F(x)$$

Then m_F is a countably additive content on $(\mathbb{R}, a(\epsilon))$
By the extension theorem, m_F extends to a measure on $(\mathbb{R}, \sigma(a(\epsilon)) = \sigma(\epsilon) = \mathcal{B}(\mathbb{R}))$

(F(x)=(x) gives Lebesgue)

Definition 1.5.4: Lebesgue-Stieltjes mesures

think of $m_F(A) = \int_A \underbrace{dF(x)}_{\text{R-S integral}}$, F is the distribution function of the measure

§1.5.1 π and λ system

Definition 1.5.5: π and λ system

Let C be a collection of sets of Ω
C is a π-system if:

- $\emptyset \in C$
- $\forall A, B \in C, A \cap B \in C$

C is a λ-system if:

- $\Omega \in C$
- if $A, B \in C,$ and $A \subseteq B,$ then $B \backslash A \in C$
- if $A_1, A_2, \cdots \in C,$ and $A_1 \subseteq A_2 \subseteq \cdots,$ then $\bigcup_{i=1}^{+\infty} A_i \in C$

Example 4

$$\epsilon = \begin{cases} (a, b] \\ (a, +\infty) \end{cases} \quad \text{is a } \pi\text{-system}$$

Exercise: if C is both a π-system and a λ-system, then it is a σ-algebra

Lemma 1.5.3: Dynkins Lemma

Let C be a π-system, then any λ-system containing C also contains the $\sigma(C)$
Hint: show that any such λ-system is also a π-system

Theorem 1.5.3: Uniqueness Theorem

Let C be a π-system, Let μ_1, μ_2 be two finite measures on $(\Omega, \sigma(C))$
Suppose that $\mu_1(A) = \mu_2(A)$ and $\mu_1(\Omega) = \mu_2(\Omega)$ on C, then $\mu_1(A) = \mu_2(A)$ on $\sigma(C)$

Proof.

$$D = \{A \in \sigma(C) : \mu_1(A) = \mu_2(A)\}$$
We know $C \in D$. wnat to show D is a λ-system
If so , by Dynkins Lemma, $\sigma(C) \subseteq D$, so that $D = \sigma(C)$

Check D is a λ-system:

- $\Omega \in D$ follows from $\mu_1(\Omega) = \mu_2(\Omega)$
- if $A, B \in D, A \subseteq B \to \mu_1(A) = \mu_2(A), \mu_1(B) = \mu_2(B)$
  $\mu_1(B \backslash A) = \mu_1(B) - \mu_1(A) = \mu_2(B) - \mu_2(A) = \mu_2(B \backslash A)$
  So, $\Rightarrow B \backslash A \in D$
- if $A_n \in D$ and A_n increasing, $\lim_{n\to+\infty} A_n = A$

$$\mu_2(A) = \mu_2 \lim_{n\to+\infty}(A_n) \underbrace{=}_{\text{cont.}} \lim_{n\to+\infty} \mu_2(A_n) = \lim_{n\to+\infty} \mu_1(A_n) = \mu_1(A)$$

■

Remark 1

Also hold for μ_1, μ_2, σ-finite

§1.6 Recitation 2 (02-14)-Exercise

EX1

Let Ω be a countable set,

$$A = \{A \subseteq \Omega : \text{ A is finite , or} A^c \text{ is finite}\}$$

① show that A is an algebra

② Let $P : A \to [0, +\infty)$ that $P(A) = \begin{cases} 0, & \text{if A is finite} \\ 1, & \text{if } A^c \text{ is finite} \end{cases}$

Is P a content/measure?
Solution: (1):

- $\emptyset \in \mathcal{A}$ because $\emptyset$ is finite
- $A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$ by defintation
- $A, B \in \mathcal{A}$
  - if A, B are finite, then $A \cup B$ is finite
  - if one of A, B is countably infinite , say B

$$(A \cup B)^c = A^c \cap B^c \text{ is finite} \Rightarrow A \cup B \in \mathcal{A}$$

(2): $\Omega = \{\omega_1, \omega_2, \cdots\}$
Let $A_i = \{\omega_i\}$ so that $P(\bigcup_i \{\omega_i\}) = P(\Omega) = +\infty$ But $\sum P(\omega_1) = 0$

EX2

Let Ω be an uncountable set, $A = \{\{\omega\}, \omega \in \Omega\}$, compute $\sigma(A)$ and justify:
Solution:

$$\sigma(A) = \{\underbrace{A \subseteq \Omega : A \text{ is countable or } A^c \text{ is countable }}_{f}\}$$

Proof.

- f is a σ-algebra
  - $\emptyset \in f$ because $\emptyset$ is countable
  - $A \in f \Rightarrow A^c \in f$ by definition
  - if $A_1, A_2, \cdots \in f$
    - if A_i are countable, then $\bigcup_{i=1}^{+\infty} A_i$ is countable
    - if one of A_i is uncountable, say A_1 thenA_1^c countable , then $(\bigcup_{i=1}^{+\infty} A_i)^c$ is uncountable
  - $\sigma(A) \subseteq f$ by $\sigma(A)$ is minimal
  - $\sigma(A) \supseteq f$
    - if $A \in f$ countable, $A = \{\omega_1, \omega_2, \cdots\} =$
    - if $A^c \in f$ countable, then

■

EX3

Let $\Omega = \mathbb{R}$

$$C_1 = \{(-\infty, b], b \in \mathbb{R}\}$$
$$C_2 = \{(a, b], -\infty \le a < b < +\infty\}$$
$$C_3 = \{(a_1, b_1] \cup (a_2, b_2] \cup (a_3, b_3] \cdots (a_n, b_n], -\infty \le a_1 < b_1 \le a_2 < \cdots < b_n < +\infty\}$$

show that $\sigma(C_1) = \sigma(C_2) = \sigma(C_3)$
Solution:

Proof. $\sigma(C_1) = \sigma(C_2)$

- $C_1 \subseteq \sigma(C_2)$ because $(-\infty, b] \in \sigma(C_2) \Rightarrow \sigma(C_1) \subseteq \sigma(C_2)$ because $\sigma(C_1)$ is minimal
- $C_2 \subseteq \sigma(C_1)$ because $(a, b] = (-\infty, b] \backslash (-\infty, a] \in \sigma(C_1) \Rightarrow \sigma(C_2) \subseteq \sigma(C_1)$ because $\sigma(C_2)$ is minimal

■

§1.6.1 Special Case:

$\Omega = \{1, 2, \cdots, N\}, F = P(\Omega), P(\{1\}) = \cdots = P\{N\} = \frac{1}{N}$
For event:$E \in \Omega, P(E) = \frac{|E|}{|\Omega|}$

Definition 1.6.1: Inclusion-Exclusion Principle

$P(E_1 \cup E_2 \cup \cdots \cup E_n) = \sum_{i=1}^{n} P(E_i) - \sum_{i<j} P(E_i \cap E_j) + \sum_{i<j<k} P(E_i \cap E_j \cap E_k) + \cdots + (-1)^{r-1} \sum_{i_1 < \cdots < i_r} P(E_{i_1} \cap \cdots \cap E_{i_r}) + \cdots + (-1)^{n-1} P(E_1 \cap E_2 \cap \cdots \cap E_n)$

Remark 1

① Prove by induction

② $P(E_1 \cup E_2 \cup \cdots \cup E_n) \leq \sum P(E_i)$
$P(E_1 \cup E_2 \cup \cdots \cup E_n) \geq \sum P(E_i) - \sum P(E_i \cup E_j)$
$P(E_1 \cup E_2 \cup \cdots \cup E_n) \leq \sum P(E_i) - \sum P(E_i \cap E_j) + \sum P(E_i \cap E_j \cap E_k)$

Example 1: Brithday Problem

N people, what is the P[at least two people have the same birthday]?
Solution:

$$\begin{aligned}
\Omega &= \{(x_1, \cdots, x_n, x_i \in \{1, \cdots, 365\})\} |\Omega| = 365^N \\
A^c &= \{(x_1, \cdots, x_N) \in \Omega : x_i \neq x_j\} |A^c| = 365 \cdot 364 \cdots (365 - N + 1) \\
P(A) &= \frac{|A^c|}{|\Omega|} = 1 \cdots (1 - \frac{1}{365}) \cdots (1 - \frac{N-1}{365})(use 1 - x \leq e^{-x}) \\
&\leq e^{-\sum_{i=0}^{N-1} \frac{i}{365}} \\
&= e^{-\frac{N(N-1)}{730}}
\end{aligned}$$

in fact if N>23, then $P(A^c) < \frac{1}{2}$

§1.7 Lecture 5 (02-17)

Useful π-system that generates B(R):

① $\mathcal{E} = \begin{cases} = \{(a, b], -\infty \leq a < & b < +\infty\} \\ = \{(a, +\infty), a & \in R\} \end{cases}$

② $\mathcal{E}_1 = \{(a, b], -\infty \leq a < b < +\infty\}$

③ $\mathcal{E}_2 = \{(a, b), -\infty \leq a < b \leq +\infty\}$

④ $\mathcal{E}_{\text{open}} = \{A \in R, A \text{ open }\}$

⑤ $\mathcal{E}_{\text{closed}} = \{A \in R, A \text{ closed }\}$

⑥ $\mathcal{E}_{\text{half}} = \{(-\infty, b], b \in \mathbb{R}\}$

easy to check π-system
Note:

- A open iff $\forall x \in A, \exists \mathcal{E}_x > 0, \text{ s.t. } (x - \mathcal{E}_2, x + \mathcal{E}_2) \subseteq A$
- A closed iff A^c open

Proof. (2)

$$\begin{aligned}
&\mathcal{E}_1 \subseteq \sigma(\mathcal{E}) \Rightarrow \sigma(\mathcal{E}_1) \subseteq \sigma(\mathcal{E}) \\
&\mathcal{E} \subseteq \sigma(\mathcal{E}_1), (a, +\infty) = \bigcup_{n\geq 1} \underbrace{(a, a+n]}_{\mathcal{E}_1} \in \sigma(\mathcal{E}_1) \\
\Rightarrow &\sigma(\mathcal{E}) \subseteq \sigma(\mathcal{E}_1)
\end{aligned}$$

(3)

$$\begin{aligned}
\mathcal{E}_1 \subseteq \sigma(\mathcal{E}_2) : (a, b] = \bigcap_{n\geq 1} \underbrace{(a, b + \frac{1}{n})}_{\mathcal{E}_2} \in \sigma(\mathcal{E}_2) \\
\mathcal{E}_2 \subseteq \sigma(\mathcal{E}_1) : (a, b) = \bigcup_{n\geq 1} \underbrace{(a, b - \frac{1}{n}]}_{\mathcal{E}_2} \in \sigma(\mathcal{E}_2)
\end{aligned}$$

(4)

$$\mathcal{E}_2 \subseteq \mathcal{E}_{\text{open}} \Rightarrow \sigma(\mathcal{E}_2) \subseteq \sigma(\mathcal{E}_{open})$$

Fact:
every open set $A \subseteq \mathbb{R}, A = \bigcup i = 1^{+\infty}(\underbrace{(x_i - \mathcal{E}_i, x_i + \mathcal{E}_i)}_{\mathcal{E}_2} \in \sigma(\mathcal{E}_2))$

$\Rightarrow \mathcal{E}_{open} \subseteq \sigma(\mathcal{E}_2) \Rightarrow \sigma(\mathcal{E}_{open}) \subseteq \sigma(\mathcal{E}_2)$
(e)

$$\begin{aligned}
&A \text{ open } \Leftrightarrow A^c \text{ closed implies} \\
&\mathcal{E}_{\text{closed}} \subseteq \sigma(\mathcal{E}_{\text{open}}) \Rightarrow \sigma(\mathcal{E}_{\text{closed}}) \subseteq \sigma(\mathcal{E}_{\text{open}}) \\
&\mathcal{E}_{\text{open}} = \sigma(\mathcal{E}_{\text{closed}})
\end{aligned}$$

■

Example 1: mismatch

N men and N hats
P[no one finds his own hats]=?
Solution:
$E_i = \{i^{th} \text{ letter in } i^{th}\text{envelope}\}$

Want to compute $P(\bigcap_{i=1}^n E_i^c) = 1 - P(\bigcup_{i=1}^n E_i)$ By inclusion-exclusion:

$$P(\bigcup_{i=1}^n E_i) = \sum_{i=1}^n P(E_i) - \sum_{i<j} P(E_i \cap E_j) + \cdots + (-1)^{n-1} P(\sum_{i=1}^n E_{i1} \cap \cdots \cap E_{ir})$$

$$P(\sum_{i=1}^n E_{i1} \cap \cdots \cap E_{ir}) = \frac{|E_{i1} \cap \cdots \cap E_{ir}|}{|\Omega|} = \frac{(n-r)!}{n!}$$

$$\sum_{i_1 < \cdots < i_r} P(E_{i1} \cap \cdots \cap E_{ir}) = \binom{n}{r} \frac{(n-r)!}{n!} = \frac{1}{r!}$$

$$P(\bigcup_{i=1}^n E_i) = 1 - \frac{1}{2!} + \frac{1}{3!} + \cdots + (-1)^{N-1} \frac{1}{N!}$$

$$P(\bigcap_{i=1}^n E_i^c) \to 1 - \frac{1}{e}$$

Exercise:

- Circle: 10 couples,
  P[no couple sit next to each other]=?
- Texas Holder:
  P[Sraight]=? P[Full House]=? sol:

$$|\Omega| = \binom{52}{5}$$

$$|Srtaight| = 10 \times (4^5 - 4)$$

$$P[Straight] = \frac{10 \times (4^5 - 4)}{\binom{52}{5}} \approxeq 0.0039$$

$$|FullHouse| = 13 \times \binom{4}{3} \times 12 \times \binom{4}{2}$$

$$P[FullHouse] = \frac{13 \times \binom{4}{3} \times 12 \times \binom{4}{2}}{\binom{52}{5}} \approxeq 0.0014$$

Conditional Probability:
"If the event B has occurred, what is the probability of event A?" $P[A\backslash B]$
N experiment: natural

$$P[A\backslash B] = \frac{\text{number of occurance of both A and B}}{\text{number of occurance of B}} = \frac{P[A \cap B]}{P[B]}$$

Definition 1.7.1

If $P[B] > 0$, then the conditional probability of A given B is $P[A\backslash B] = \frac{P[A \cap B]}{P[B]}$.

Example 2

① 2 kids

- P[two boys\at least one boy] =?

- P[two boys\younger kid is boy] =? 1/2
- P[two boys\at least one boy born on Tuesday] =?

(1)

$$A = \{BB\}, B = \{BG, GB, BB\}$$
$$P[A \backslash B] = \frac{P[A \cap B]}{P[B]} = \frac{P[A]}{P[B]} = \frac{1/4}{3/4} = \frac{1}{3}$$

(3)

$$\begin{aligned}
\Omega &= \{B_iB_j, B_iG_j, G_iB_j, G_iG_j, i, j = 1, \cdots, 7\} \\
A &= \{B_iB_j, i, j = 1, \cdots, 7\} \\
B &= \{B_2B_j, B_iB_2, B_2G_j, i, j = 1, \cdots, 7\} \text{ 13+14 elements} \\
A \cap B &= \{B_2B_j, B_iB_2, i, j = 1, \cdots, 7\} \text{ 13 elements} \\
P[A \backslash B] &= \frac{13}{27}
\end{aligned}$$

§1.8 Lecture 6 (02-19)-Conditional Probability

Definition 1.8.1: Law of total probability

$$P[A|B] = \frac{P[A \cap B]}{P[B]}$$
$$P[A] = P[A|B] \cdot P[B] + P[A|B^c]P[B^c]$$

More generally, Let $(B_i)_{i=1}^n$ be a partition of Ω (i.e. $B_i \cap B_j = \emptyset$ and $\bigcup_{i=1}^n B_i = \Omega$), then:

$$P[A] = \sum_{i=1}^{n} P[A|B_i]P[B_i]$$

Proof:

$$P[A] = P[A \cap B] + P[A \cap B^c] = P[A|B]P[B] + P[A|B^c]P[B^c]$$

Example 1

① draw balls randomly from box A(3B2W) to B(4B3W) ,then draw from B randomly,P[The second draw is Black]
By Law of total probability: P=P[2nd Black|1st Black]P[1st Black]+P[2nd Black|1st White]P[1st White]
$=\frac{5}{8} \times \frac{3}{5} + \frac{1}{2} \times \frac{2}{5} = \frac{23}{40}$

Reverse Question:

If A happens, which B_i is the most likely? **Bayes' Formula:**

$$P[B_i|A] = \frac{P[B_i \cap A]}{P[A]} = \frac{P[A|B_i]P[B_i]}{\sum_{i=1}^{n} P[A|B_i]P[B_i]}$$

Example 2

① Covid Test:

- false negative:P[negative|infected]=0.05
- false positive:P[positive|not infected]=0.01

Suppose 5% of the population are infected, then P[infected|positive]=?

Solution.

$$P[V|P] = \frac{P[P|V]P[V]}{P[P|V]P[V] + P[P|V^c]P[V^c]} = \frac{5}{6}$$

② Prisoner parados:A,B,C. 2 executed, 1 pardoned.
A asked :"Please tell me the name of someone else who will be executed"
Guard:"B will be executed"
P[A survive|B will be executed]=?

Solution.

$$\begin{aligned} P[A|\text{Guard says B}] &= \frac{P[\text{Guard says B}|A]P[A]}{P[\text{Guard says B}|A]P[A] + P[\text{Guard says B}|B]P[B] + P[\text{Guard says B}|C]P[C]} \\ &= \frac{\frac{1}{6}}{\frac{1}{6} + 0 + \frac{1}{3}} = \frac{1}{3} \end{aligned}$$

Solution.

$$\begin{aligned} \Omega &= \{(\text{survive person,name mentioned by Guard})\} \\ &= \begin{cases} (A, B), & \frac{1}{6} \\ (A, C), & \frac{1}{6} \\ (B, C), & \frac{1}{3} \\ (C, B), & \frac{1}{3} \end{cases} P[A|B] = \frac{P(A, B)}{P(B)} = \frac{\frac{1}{6}}{\frac{1}{6} + \frac{1}{3}} = \frac{1}{3} \end{aligned}$$

③ Two envelope problem:X,2X ,switch or not?

-

§1.9 Lecture 7 (02-21)-Independence

Definition 1.9.1: Independence

$$P[A|B] = P[A]$$

then we say A and B are independent
Two events A and B are independent iff

$$P[A \cap B] = P[A]P[B]$$

(A,B independent $\Rightarrow A^c, B$ independent)

Definition 1.9.2: Multiple events Independence

The events $A_1, A_2, \cdots, A_n$ are independent iff

$$P[A_1 \cap A_2 \cap \cdots \cap A_n] = P[A_1]P[A_2]\cdots P[A_n]$$

Notice: Stronger than the condition$P[A \cap B] = P[A]P[B]$ (pairwise independence)

Example 1: pairwise independence<Independence

x_1, x_2, x_3 are coin flips:

$$P(x_i = 1) = P(x_i = 0) = \frac{1}{2}$$
$$A_1 = \{x_2 = x_3\}, A_2 = \{x_1 = x_3\}, A_3 = \{x_1 = x_2\}$$
$$P[A_i] = \frac{1}{2}, P[A_i \cap A_j] = P[x_1 = x_2 = x_3] = \frac{1}{4} = P[A_i]P[A_j] \text{ pairwise independent}$$

$P[A_1 \cap A_2 \cap A_3] = P[x_1 = x_2 = x_3] = \frac{1}{4}$ not independent

Example 2: Independence

Independent trials, each with success probability p, fail probability 1-p
$P_{n,m}[\underbrace{\text{first n success}}_{A_{n,m}}$ occured before first m failures]=?

Solution. (Pascal)
By Law of total probability:

$$\begin{aligned} P[A_{n,m}] &= P[A_{n,m}| \text{ 1st success }]P[\text{1st success}] + P[A_{n,m}| \text{ 1st failure }]P[\text{1st failure}] \\ &= P[A_{n-1,m}] \cdot P + P[A_{n,m-1}] \cdot (1-P) \\ \Rightarrow P_{n,m} &= P_{n-1,m} \cdot P + P_{n,m-1} \cdot (1-P) \end{aligned}$$

Boundary condition:$P_{0,m} = (1-P)^m, P_{n,0} = 0$

Or building generating function
(Fermat):
{First n success before first m failure}
$\Leftrightarrow$ {at least n success in the first m+n-1 trials} (ex)
P[exact k success in n+m-1 trials] $= \binom{n+m-1}{k} p^k (1-p)^{m+n-1-k}$ Binomial distribution
$\Rightarrow P$[at least n success in the first m+n-1 trials] $= \sum_{k=n}^{m+n-1} \binom{n+m-1}{k} p^k (1-p)^{m+n-1-k}$

Example 3

Multiple choice test, m options, p-knows the answer, 1-p random guess
P[knows the answer|correct]=$\frac{p}{p+(1-p)(\frac{1}{m})} = \frac{mp}{mp+1-p}$ Bayes

Example 4

Gambler's ruin:
bet 1 dollar each time, p-win, 1-p-lose, initial amount of money=$i \in [0, N]$
P_i[Reach N before reaching 0] = win times=N-i+lose times, Method2: $p_i = pp_{i+1} + (1-p)p_{i-1}$
characteristic polynomials: take $p_i = cr^i$

$$\begin{aligned} pr^2 - r + (1-p) = 0 \\ r = 1, \frac{1-p}{p} \\ \text{if } p \neq 1-p, \text{ then } p_i = c_1 + c_2(\frac{1-p}{p})^i \\ \text{if } p = 1-p, \text{ then } p_i = c_1 + c_2 i, c_1, c_2 \text{ determined by } p_0, p_N \end{aligned}$$

One dimentional random walk:

$$S_n = S_0 + X_1 + X_2 + \cdots + X_n$$
$$X_i \text{ are i.i.d. } P(X_i = 1) = P, P(X_i = -1) = 1 - P$$

Example 5

1. secretary problem:
   N candidates
   After each interview, immediately make offer or rejection
   what is the best strategy
   maximize P[best candidate is offered]

   Solution. Not making offer to first r candidate, make an offer to the next candidate that is better than {1,2,...r}
   P[Best candidate is offered]
   $= \sum_{i=0}^{N} \underbrace{P[\text{best candidate is i}]}_{\frac{1}{N}} P[\text{Best candidate is offerd|Best candidate=i}]$

§1.10 Lecture 7 (02-24)-Random walk

Definition 1.10.1: 1-dimentional random wlk

$S_n = S_0 + x_1 + x_2 + \cdots + x_n$ (x_i) i.d.d(independent+identically distributed)

$P(x_i = 1) = p, P(x_i = -1) = 1 - p$

Generalization: 2-D(simple) random walk $S_n = S_0 + x_1 + x_2 + \cdots + x_n$ (x_i) i.d.d

$$P(x_i = \pm e_1) = P(x_i = \pm e_2) = \frac{1}{4}$$

P[random walk(starting at (i,j)) exit the boundary through A]=?
By conditioning, $P_{i,j} = \frac{1}{4}P_{i+1,j}\frac{1}{4}P_{i-1,j} + \frac{1}{4}P_{i,j-1} + \frac{1}{4}P_{i,j+1}$
Boundary condition: $P_{i,j} = 1$ if $(i,j) \in A$, and $P_{i,j} = 0$ otherwise
We have

$$\begin{aligned} P_{i,j} &= \frac{1}{4}P_{i+1,j}\frac{1}{4}P_{i-1,j} + \frac{1}{4}P_{i,j-1} + \frac{1}{4}P_{i,j+1} \\ &\Leftrightarrow \frac{1}{4}(P_{i+1,j} + P_{i-1,j} - 2P_{i,j}) + \frac{1}{4}(P_{i,j-1} + P_{i,j+1} - 2P_{i,j}) \quad = 0 \\ \Leftrightarrow \Delta P &= 0 \end{aligned}$$

Discrete Laplacian on Z^2,

$$\Delta P(x) = \sum_{y \approx x} \frac{1}{4}(P(y) - P(x))$$

§1.10.1 Random Variable and Measurable functions

Example 1

① 2 coinflips, $\Omega = \{HH, TT, HT, TH\}, f = P(\Omega)$ X= #heads,then we may write $X : \Omega \to \mathbb{N}, X = 2 \cdot 1_{HH} + 1_{HT} + 1_{TH}$

$$P(X \leq x) = P(\{\omega \in \Omega : X(\omega \leq x)\})$$

It is called the distribution function of X (right continuous increasing function)

Definition 1.10.2

Let $(\Omega_1, F_1), (\Omega_2, F_2)$ be two measurable space

- A map $X : \Omega_1 \to \Omega_2$ is called measurable iff $\forall A \in F_2, X^{-1} \in F_1$, $X^{-1} = \{\omega \in \Omega_1 : X(\omega) \in A\}$
- if (Ω_1, F_1, P) is a probability space,then a measurable funtion $X : \Omega_1 \to \Omega_2$ is called a random variable
- if $(\Omega_2, F_2) = (R, B(R))$ then a measurable function is called a Borel funtion
- if $X : (\Omega_1, F_1, P) \to (R, B(R))$ then X is a R-valued random variable and

$$F_X(x) = P(X^{-1}(-\infty, x]) = P(X \leq x)$$

is called the distribution function of X

Remark: By (a), we know that for any $B \in B(R)$ we can define $P(X^{-1}(B))$

Example 2

① Let (Ω, F, P) be a probability space,$A \in F$

Then $1_A : \Omega \to \{0,1\}$ is a random variable, $1_A(\omega) = \begin{cases} 1, & \text{if } \omega \in A \\ 0, & \text{if } \omega \in A^c \end{cases}$ We need to show $\forall B \in B(R), 1_A^{-1}(B) \in F$

In fact, $1_A^{-1}(B) = \begin{cases} \emptyset, & \text{if } 0,1 \notin B \\ A, & \text{if } 1 \in B, 0 \notin B \\ A^c, & \text{if } 0 \in B, 1 \notin B \\ \Omega, & \text{if } 0,1 \in B \end{cases}$

In practice, only need to check the pre-image on a smaller set:

Proposition 1.10.0

Let $(\Omega_1, F_1), (\Omega_2, F_2)$ be two measurable space
$\mathcal{E} \subseteq F_2$ such that $\sigma(\mathcal{E}) = F_2$
Then $X : \Omega_1 \to \Omega_2$ is measurable if $\forall A \in \mathcal{E}, X^{-1}(A) \in F_2$

Proof. Let $G = \{B \subseteq \Omega_2, X^{-1}(B) \in F_1\}$
Then G is a σ-algebra (ex)
Therefore if $G \supseteq \mathcal{E}$ then $G \supseteq \sigma(\mathcal{E}) = F_2$ ■

Corollary 1.10.1

et(Ω, F) be a measurable space, then $X : \Omega \to R$ is a Borel function iff the following are true.

- $\{x < a\} \in F$ for $\forall a \in R$
- $\{x \leq a\} \in F$ for $\forall a \in R$
- $\{x > a\} \in F$ for $\forall a \in R$
- $\{x \geq a\} \in F$ for $\forall a \in R$

Proof. $\{x < a, a \in \mathbb{R}\} = \{X^{-1}(-\infty, a), a \in \mathbb{R}\}$
suffices to show:$\mathcal{E} = \{(-\infty, a), a \in \mathbb{R}\}$
satisfies $\sigma(\mathcal{E}) = B(\mathbb{R})$
indeed:

$$[a, +\infty) = (-\infty, a)^c \in \sigma(\mathcal{E}) \tag{1.1}$$
$$[a, b) = [a, +\infty)\backslash(b, +\infty) = (-\infty, a)^c\backslash(-\infty, b)^c \in \sigma(\mathcal{E}) \tag{1.2}$$
$$\Rightarrow \sigma(\mathcal{E}) = B(R) \tag{1.3}$$

■

Remark 1

a R-valued random variable is a function $X : \Omega \to \mathbb{R}$ such that $\forall a \in \mathbb{R}, \{X < a\} \in F$

Lemma 1.10.1

A distribution funtion F satisfies:

① $\lim_{x\to-\infty} F(x) = 0, \lim_{x\to+\infty} F(x) = 1$

② $F(x) \leq F(y)$ if $x \leq y$

③ F is right continuous,$F(x+h) \to F(x)$ as h decreases to 0

Proof. $F(x) = P(X \leq x)$ (2) is immediate
(1): Let $A_n = \{x \leq -n\}$ then by continuous of measure:$lim_{n\to+\infty} = \lim_{n\to+\infty} = P(\bigcap_{n\geq 1} A_n) = 0$
Then by monotonicity of F, $lim_{x\to-\infty}F(x) = 0$
(3): Let $B_n = \{X \leq x + \frac{1}{n}\}$(decreaing),
then $F(x+\frac{1}{n}) = P(B_n) \underbrace{\to}_{n\to+\infty, cont.} P(\bigcap_{n\geq 1} B_n) = P(X \leq x) = F(x)$
again use F increasing to conclude ■

Lemma 1.10.2

① $F(y) - F(x) = P(x < X \leq y), \forall x \leq y$

② $F(x) - \lim_{h\to+\infty} F(x-h) = P(X = x)$

③ $B_n = \{x - \frac{1}{n} < X \leq x\}$(decreaing),Then $\bigcap_{n\geq 1} B_n = \{X = x\}$
$\Rightarrow F(x) - F(x-\frac{1}{n}) = P(B_n) \underbrace{\to}_{n\to+\infty} P(X = x)$
again use F increasing to conclude

Naturally:

- (absolute) continuous random variable
- discrete random variable

§1.11 Lecture 8 (02-26)

properties of measurable functions(R.V.s):

- Let $\{X < Y\} = \{\omega \in \Omega_1 : X(\omega) < Y(\omega)\}$
  $\{X > Y\} = \{\omega \in \Omega_1 : X(\omega) > Y(\omega)\}$

Lemma 1.11.1

Let (Ω, F) be a measurable space, X, Y are Borel Functions

- $\{X < Y\}, \{X \leq Y\}, \{X = Y\}, \{X \neq Y\} \in F$
- $X + Y, X \cdot Y, X/Y$ are Borel functions

Proof. (1):Use that Q is dense in R

$$\{X < Y\} = \bigcup_{q \in Q} \{X < q < Y\} = \bigcup_{q \in Q} \underbrace{\{X < q\}}_{\in F} \cap \underbrace{\{q < Y\}}_{\in F} \in F$$

$$\{X = Y\} = \bigcap_{n \geq 1} \{X < Y + \frac{1}{n}\} \cap \bigcap_{n \geq 1} \{X > Y - \frac{1}{n}\} \in F$$

(2):Fact: if Y is Borel, then aY+b,$a, b \in R$ is Borel (ex)
Then

$$\forall a \in \mathbb{R}, \{X + Y < a\} = \{X < a - Y\} \in F \text{ by}(1)$$

so $X + Y$ is Borel

$$\{X^2 < a\} = \begin{cases} \emptyset, \\ \{X < \sqrt{a}\} \cap \{X > -\sqrt{a}\} \in F \end{cases}$$

so X^2 is Borel
$X \cdot Y = \frac{1}{4}[(X + Y)^2 - (X - Y)^2]$ is Borel ∎

Lemma 1.11.2

Let $(X_n)_{n \geq 1}$ be a sequence of Borel functions on (Ω, F),
Then the following are Borel functions:

- $\sup_{n \geq 1} X_n, \inf_{n \geq 1} X_n, \limsup_{n \geq 1} X_n, \liminf_{n \geq 1} X_n$

In particular, if $\lim_{n \to +\infty} X_n$ exists, then $\lim_{n \to +\infty} X_n$ is Borel

Proof.

$$\{\sup_{n \geq 1} X_n < a\} = \bigcap_{n \geq 1} \{X_n < a\} \in F \forall a \in \mathbb{R} \Rightarrow \sup_{n \geq 1} X_n \text{ is a Borel function}$$

$$\{\inf_{n \geq 1} X_n > a\} = \bigcap_{n \geq 1} \{X_> < a\} \in F \forall a \in \mathbb{R} \Rightarrow \inf_{n \geq 1} X_n \text{ is a Borel function}$$

$$\limsup X_n = \inf_{m \geq 1} \sup_{n \geq m} X_n \text{ is a Borel function}$$

$$\liminf X_n = \sup_{m \geq 1} \inf_{n \geq m} X_n \text{ is a Borel function}$$

∎

Lemma 1.11.3

Let $(\Omega_1, F_1), (\Omega_2, F_2), (\Omega_3, F_3)$ be two measurable space
$X : \Omega_1 \to \Omega_2, Y : \Omega_2 \to \Omega_3$ are measurable, then $Y \circ X : \Omega_1 \to \Omega_3$ is measurable

Proof.

$$\forall A \in F_3, (Y \circ X)^{-1}(A) = X^{-1}(\underbrace{Y^{-1}(A)}_{\in F_2}) \in F_1$$

∎

Definition 1.11.1: σ-algebra generated by r.v.

If $X : \Omega \to \mathbb{R}$ is a random variable, then

$$\sigma(x) = \{X^{-1}(A), A \in B(\mathbb{R})\}$$

is called the σ-algebra generated by X
Let $(X_i)_{i\in I}$ be a family of r.v.s

$$\sigma(X_i, i \in I) = \sigma(\bigcup_{i\in I} \sigma(X_i))$$

is the σ-algebra generated by $(X_i)_{i\in I}$

Remark 1

$\sigma(X)$ is the smallest σ-algebra such that X is measurable

Example 1

① (Ω, F, P) ,Let $X : \Omega \to \mathbb{R}$ be a random variable, $X = b_1 1_{A_1} + b_2 1_{A_2} + \cdots b_n 1_{A_n}, b_i \in \mathbb{R}, A_i \in F, A_i \cap A_j = \emptyset$
if $b_j \neq b_j$
Then $\sigma(X) = \sigma\{A_1, A_2, \cdots, A_n\} =$"finite disjioint union of $A_1 \cdots A_n$"

Proof.
- $\supseteq$: Note that $X^{-1}(\{b_1\}) = A_1, \cdots, X^{-1}(\{b_n\}) = A_n$
$\Rightarrow A_1, \cdots, A_n \in \sigma(x), \sigma(A_1, \cdots, A_n) \in \sigma(x)$
- $\subseteq$

Lemma 1.11.4

$\sigma(X) = \sigma(\{X \leq a\}, a \in \mathbb{R})$

it suffices to show $\forall a \in \mathbb{R}, \{X \leq a\} \in \sigma(\{A_1, \cdots, A_n\})$
$\{X \leq a\} =$"finite disjioint union of A_i and $(\bigcup A_i)^c$"

■

Two specific cases: Discrete and (Absolutely)continuous random variables

Definition 1.11.2

A r.v. is discrete if it takes values in a countable set $\{X_1, X_2, \cdots\}$
prob mass function:$f(x) = P(X = x)$

Remark 2

We say $x_1, x_2, \cdots$, are atoms of F_x

Definition 1.11.3

A r.v. is (absolutely) continuous if

$$F_x(x) = P(X \leq x) = \int_{-\infty}^{x} f(u)du$$

for some integrable function $f : R \to [0, +\infty), f(x) = F_x'(x)$ is called the probability density function of X

Remark 3

F_x is absolutely continuous
F is absolutely continuous iff

$$\forall \epsilon > 0, \exists \delta > 0$$

s.t. for any finite collecion of intervalsa_i, b_i s.t.

$$\sum_{i=1}^{n} |b_i - a_i| < \delta \Rightarrow \sum_{i=1}^{n} |F(b_i) - F(a_i)| < \epsilon$$

$$C^1 \Rightarrow \text{absolutely continuous} \Rightarrow \text{uniformly continuous}$$

Remark 4

X is a singular Continuous R.V. if F_x is continuous but F_x is not absolutely continuous (ex: Cantor funtion)

Discrete R.V.:

$$F_x(x) = P(X \leq x) = \sum_{x_i \leq x} \underbrace{f(x_i)}_{\text{prob mass function}}$$

$$f(x) = F_x(x) - \lim y \nearrow x - F_x(y)$$

Definition 1.11.4

The expectation/mean of a discrete r.v. with prob mass function f is

$$E[X] = \sum_{x:f(x)>0} x_i f(x_i)$$

whenever the sum is absolutely convergent

Remark 5

absolutely convergent $\Rightarrow$ order of the sum does not matter

Example 2

① 2 coin flips,X=#heads, $f(0) = f(2) = \frac{1}{4}, f(1) = \frac{1}{2}$
$E[X] = 0 \cdot \frac{1}{4} + 1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{4} = 1$

Lemma 1.11.5

Let X be a r.v. taking values in N.
Then

$$E[X] = \sum_{n=1}^{+\infty} P(X \geq n)$$

Proof.

$$\begin{aligned} E[x] &= 1 \cdot P(X=1) + 2 \cdot P(X=2) + \cdots \\ &= P(X=1) + P(X=2) + P(X=3) + \cdots \\ & \qquad\qquad\quad P(X=2) + P(X=3) + \cdots \\ & \qquad\qquad\qquad\qquad\quad\; P(X=3) + \cdots \end{aligned}$$

$P(X \geq n)$ ■

§1.12 Lecture 9 (02-28)-Expectation

EX1

we have offers $X_1, X_2, \cdots$ continuous R.V. i.i.d
$T = inf\{n > 1, X_n > X_1\}$ Compute $E[T]$
Solution.

$$E[T] = \sum_{n=2}^{+\infty} nP(T=n)$$

By lemma,

$$\begin{aligned} E[T] &= \sum_{n\geq 1} P(T \geq n) = \sum_{n \geq 1} P(X_1 \geq X_2 \geq \cdots \geq X_n) = \frac{1}{n-1}, n \geq 2 \\ &= +\infty \end{aligned}$$

EX2

Independent trials, each with success probability p, fail probability 1-p
Compute P[There is n consecutive successes before m consecutive failures] Condition on the first trial:

$$P[A] = p \cdot P[A|H] + (1-p) \cdot P[A|T]$$

Multiple conditioning: $P(E_1 \cap E_2 \cdots E_n) = P(E_1)P(E_2|E_1) \cdots P(E_n|E_1 \cap \cdots E_n)$
Let $B = \{HH \cdots H\}$ (2nd–n-th)

$$P(A|H) = P(B)P(A|H \cap B) + P(B^c)P(A|H \cap B^c) = p^{n-1} \cdot 1 + (1 - p^{n-1}) \cdot P(A|T)$$

Let $C = \{TT \cdots T\}$ (2nd–m-th)

$$P(A|T) = P(C)P(A|T \cap C) + P(C^c)P(A|T \cap C^c)$$

Let q=1-p,

$$\begin{aligned}
P(A|H) &= \frac{p^{n-1}}{p^{n-1} + q^{m-1} - p^{n-1}q^{m-1}} \\
P(A|T) &= \frac{(1-q^{m-1})p^{n-1}}{p^{n-1} + q^{m-1} - p^{n-1}q^{m-1}} \\
P(A) &= \frac{p^{n-1}(1-q^{m-1})}{p^{n-1} + q^{m-1} - p^{n-1}q^{m-1}}
\end{aligned}$$

§1.13 Lecture 10 (03-02)-Var

Lemma 1.13.1: change of variable

Let $g : R \to R$ X be a r.v. with probability mass function f ,then

$$E[g(x)] = \sum_{x:f(x)>0} g(x)f(x)$$

Proof.

$$\begin{aligned}
E[g(x)] &= \sum_y y \cdot P(g(x) = y) = \sum_y \sum_{x:g(x)=y} y \cdot P(X = x) \\
&= \sum_x g(x)P(X = x) = \sum_x g(x)f(x)
\end{aligned}$$

■

Definition 1.13.1

Let $k \in N$,the k-th moment of X is $mk := E[X^k]$ as long as the expectation exists
Let $k \in N$,the k-th central moment of X is $\sigma k := E[X - E[X]]^k$ as long as the expectation exists

- $\sigma_2 = Var(X) = E(X - E[X])^2$ varience,"deviation fluctuation" from the mean

- $\sigma = \sqrt{Var(X)}$ standard deviation

Fact:

- $E[aX + bY] = aE[x] + bE[Y] a, b \in R$
  $\to Var(X) = E(X - E[x])^2 = E[X^2] - (E[X])^2$

Example 1

① Bernoulli(p) $P(X = 1) = p, P(X = 0) = 1 - p$
$E[X] = 1 \cdot P(X = 1) = p, Var(X) = p - p^2$

② binomial(n,p) $P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$

$$E[X] = \sum_{k=0}^{n} kP(x = k) = \sum_{k} k \cdot \binom{n}{k} p^k (1-p)^{n-k}$$

Recall

$$(1 + x)^n = \sum_{k=0}^{n} x^k \cdot \binom{n}{k}$$

,differentiate both sides w.r.t x, then

$$n(1 + x)^{n-1} = \sum_{k=1}^{n} k \cdot x^{k-1} \cdot \binom{n}{k}$$

Let $q = 1 - p, x = \frac{p}{q}$, then

$$\begin{aligned} E[X] &= \sum_{k=1}^{n} k \cdot \binom{n}{k} (\frac{p}{q})^{k-1} \cdot q^n \cdot \frac{p}{q} \\ &= n \cdot \frac{p}{1-p} \cdot (1 + \frac{p}{1-p})^{n-1} = np \\ Var(X) &= np(1-p) \end{aligned}$$

sol n': we have $X = Y_1 + Y_2 + \cdots + Y_n$ such that $Y_i \sim Bernoulli(p)$ and Y_i independent

$$E[X] = \sum_{i=1}^{n} E[Y_i] = np$$

$$\begin{aligned} Var(X) &= E[X^2] - (E[X])^2 = E(\sum_{i=1}^{n} Y_i)^2 - (np)^2 \\ &= \sum_{i=1}^{n} E[Y_i]^2 - 2\sum_{i<j}^{n} E[Y_i]E[Y_j] - (np)^2 \\ &= np + 2p^2 \frac{n(n-1)}{2} - (np)^2 \\ &= np(1-p) \end{aligned}$$

If x,y are independent, then $E[XY] = E[X]E[Y]$
Proof:

$$E[XY] = \sum_{x,y} xyP(X = x, Y = y) = \sum_{x,y} xyP(X = x)P(Y = y)$$
$$= \sum_{x} xP(X = x) \sum_{y} yP(Y = y) = E[X]E[Y]$$

Definition 1.13.2

If E[XY]=E[X]E[Y], then we say X,Y are uncorrelated
Covariance

$$Cov(X, Y) = E[XY] - E[X]E[Y] = E[(X - E[X])(Y - E[Y])]$$

Correlation

$$\rho(X, Y) = \frac{Cov(X, Y)}{\sqrt{Var(X)Var(Y)}}$$

Lemma 1.13.2

$|\rho(X, Y)| \le 1, \rho(X, Y) = \pm 1$ iff $Y = aX + b$ for some $a, b \in \mathbb{R}$

Proof.

$$\rho(X, Y) = \frac{Cov(X, Y)}{\sqrt{Var(X)Var(Y)}} = \frac{E[(X - E[X])(Y - E[Y])]}{\sqrt{Var(X)Var(Y)}}$$

by Cauchy-Schwarz inequality: $|\rho(X, Y)| \le 1$

■

Cauchy-Schwarz inequality:

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

Proof.

Since $0 \le E[(aX - bY)^2] \forall a, b \in R \Rightarrow a^2E[X^2] + 2abE[XY] + b^2E[Y^2] \ge 0$

$\Rightarrow \frac{1}{4}\Delta = E[XY]^2 - E[X^2]E[Y^2] \le 0$

■

Example 2

① $X \sim Geometric(p), P(X = k) = (1 - p)^{k-1} \cdot p$
$E[X] = \frac{1}{p}, Var(x) = \frac{1-p}{p^2}$

§1.14 Lecture 11 (03-05)-poisson random variable

Poisson random variable:
Observe the number of customers in the past days, X_i is the number of customers on day i
How to predict the number of customers tomorrow?

- One may take $\bar{X_n} = \frac{1}{n}\sum_{i=1}^{n} X_i$
- n intervals during each interval, at most 1 customers
  $\Rightarrow$ number of customers in n intervals is $Bern(p)$
  Take P, s.t. $np = E[X] = \lambda$
  number of customer is Binomial$(n, \frac{\lambda}{n})$

$$
\begin{aligned}
P(X = k) &= \binom{n}{k}\left(\frac{\lambda}{n}\right)^k (1 - \frac{\lambda}{n})^{n-k} \\
&= \frac{n!}{(n-k)!n^k}\frac{\lambda^k}{k!}(1 - \frac{\lambda}{n})^{n-k} \\
&\Rightarrow (n \to +\infty, k \text{ fixed})\frac{\lambda^k}{k!}e^{-\lambda}
\end{aligned}
$$

Definition 1.14.1

X is a Poisson(x) is given by probability mass function

$$f(k) = \frac{\lambda^k}{k!}e^{-\lambda}, k \in N$$

if $X \sim Poisson(\lambda)$, then

$$E[X] = \sum_{k\in N^+} k \cdot \frac{\lambda^k}{k!}e^{-\lambda} = \lambda\sum_{k=0}^{+\infty}\frac{\lambda^k}{k!}e^{-\lambda} = \lambda e^{\lambda-\lambda} = \lambda$$

Continuous r.v.

$$F_x(x) = P(X \leq x) = \int_{-\infty}^{x} f(u)du, f(x) = F_x'(x)$$

is the probability density function of X

$$P(x \leq X \leq X + dx) = \int_{x}^{x+dx} f(u)du \approx f(x)dx$$

Expectation: the expectation of a r.v. X is defined by

$$E[X] = \int_{-\infty}^{+\infty} xf(x)dx = \int_{-\infty}^{+\infty} xdF_x(x)$$

k-th moment:

$$E[x^k] = \sum_{-\infty}^{+\infty} x^k f(x)dx$$

$$Var[x] = E[(X - E[X])^2] = \int_{-\infty}^{+\infty} (x - E[X])^2 f(x)dx$$

Recall if X is a N-valued r.v. then

$$E[x] = \sum_{n \in N} P(x \geq n)$$

Lemma 1.14.1

if X is a non-negative r,v, with density function f then

$$E[X] = \int_0^{+\infty} P(X > x)dx = \int_0^{+\infty} (1 - F_x(x))dx$$

Proof.

$$\begin{aligned}\int_0^{+\infty} P(X > x)dx &= \int_0^{+\infty} \int_x^{+\infty} f(y)dydx \\ &= \int_0^{+\infty} \int_0^{y} dx f(y)dy = \int_0^{+\infty} yf(y)fy = E[X]\end{aligned}$$

■

How to define $\int xdF_x(x)$ in general?

Definition 1.14.2: Lebesgue integral and expectation

recall Riemann integral

$$\sum f(x_i^\star)\Delta x_i \to \int f(x)dx$$

make sense if f has finitely many discontinuities (for $f = 1_Q$ R-I does not exist)
Lebesgue integral:

$$\{x \in \mathbb{R} : f(x) \in [y_i, y_i + \Delta y_i)\} = f^{-1}([y_i, y_i + \Delta y_i))$$

$$\text{idea: } \sum m(f^{-1}([y_i, y_i + \Delta y_i))) \cdot \Delta y_i \to \int f(x)dx$$

$$\int 1_Q(x)dx = 0$$

More generally, given any measure space $(\Omega, \mathcal{F}, \mu)$ and Borel function f
Define $\int_\Omega fdu$

Step1: If $f = 1_A$,where $A \in F$ define $\int_\Omega f d\mu = \int_\Omega 1_A d\mu = \mu(A)$
Step2: simple functions:

$$f = \sum_{i=1}^{n} a_i 1_{A_i}, A_i \in F \text{ and } a_i \geq 0, A_i \cap A_j = \emptyset$$

$$\text{Define } \int_\Omega f d\mu = \sum_{i=1}^{n} a_i \mu(A_i)$$

Fact: if f,g are simple function then f+g,fg,max{f,g},min{f,g} are simple functions

Proposition 1.14.0

if f,g are simple function, then

- $\int_\Omega a f d\mu = a \int_\Omega f d\mu \forall a \in \mathbb{R}$
- $\int_\Omega (f+g) d\mu = \int_\Omega f d\mu + \int_\Omega g d\mu$
- if $f \leq g$ then $\int_\Omega f d\mu \leq \int_\Omega g d\mu$

Step3: approximates non-negative Borel functions by simple functions
Let $f \geq 0$ Borel, then $f = \sup_i f_i$,where

$$f_i = \sum_{k=0}^{i \cdot 2^i} \frac{k-1}{2^i} 1_{\{\frac{k-1}{2^i} \leq f < \frac{k}{2^i}\}} + i 1_{\{f > i\}}$$

$f_i \geq 0$ are simple and $\lim_{i \to \infty} f_i = \sup_i f_i = f$

Definition 1.14.3

For every non-negative Borel function f, define

$$\int_\Omega f d\mu = \sup_i \int_\Omega f_i d\mu$$

Q:if $f = \sup f_i, f = \sup g_i$ Does $\sup_i \int f_i d\mu = \sup_i \int g_i d\mu$?

Consistency follows from:
Monotone Convergence theorem: For every increasing sequence $\{f_n\}$ of measurable functions:

$$\limsup_n \int_\Omega f_n d\mu = \int_\Omega \limsup_n f_n d\mu$$

(If (X_n)) is a sequence of r.v.s, $X_n \nearrow x$ then $\lim_{n \to +\infty} E[X_n] = E[\lim_n X_n] = E[X]$
Assume MCT:

$$\text{If } f = \sup f_i = \sup g_i$$
$$\text{Then } g_i \leq \sup_i f_i$$

§1.15 Recitation (03-07)

Problem 1

Let $X \sim Poisson(\lambda_1), Y \sim Poisson(\lambda_2), independent$
Compute the probability mass function of $X + Y$

$$\begin{aligned}
P(X+Y=n) &= \sum_{k=0}^{n} P(X=k)P(Y=n-k) \\
&= \sum_{k=0}^{n} \frac{\lambda_1^k}{k!} e^{-\lambda_1} \cdot \frac{\lambda_2^{n-k}}{(n-k)!} e^{-\lambda_2} \\
&= \frac{e^{-(\lambda_1+\lambda_2)}}{n!} \sum_{k=0}^{n} \binom{n}{k} \lambda_1^k \lambda_2^{n-k} \\
&= \frac{e^{-(\lambda_1+\lambda_2)}}{n!} (\lambda_1+\lambda_2)^n
\end{aligned}$$

$\sum_k f_x(k) f_Y(n-k) = f_X * f_Y$

Problem 2

m balls, n boxes, uniform random
Compute the E[number of empty boxes]
number of empty boxes:$1_{b_1 empty} + \cdots 1_{b_n empty}$
linearity of expectation:

$$\begin{aligned}
E[\#] &= \sum_{i=1}^{n} E 1_{b_i empty} \\
&= nP(\text{box 1 is empty}) \\
&= n(\frac{n-1}{n})^m
\end{aligned}$$

Note:

$$E[X] = \int X dP$$
$$E[X] = \int 1_A dP = P(A)$$

Problem 3

Coupon collector: n types of coupons, pick one at random
T_n =time to complete the collection of n coupons
Compute $E[T_n]$ Let T_i =time to collect the i-th new coupon
$T_n = T_1 + T_2 + \cdots + T_n$
$E[T_1] = 1, T_j - T_{j-1} \sim Geo(1 - \frac{i-1}{n})$ every trial with success prob:$1 - \frac{i-1}{n}$

$$E[T_j - T_{j-1}] = \frac{1}{1 - \frac{i-1}{n}}$$
$$E[T_n] = \sum_{j=1}^{n} E[T_j - T_{j-1}] = n \sum_{j=1}^{n} \frac{1}{j}$$

Problem 4

Let X be a r.v., $E[X] = 1$
show that

$$t \in (0,1), P(X > t) > \frac{(1-t)^2}{E[X^2]}$$

(hint:Cauchy Schwarz)

$$\text{Let } Y = 1_{X>t}$$
$$E[\mathbf{1}_{X>t}] = P(X > t) = E[Y] = E[Y^2]$$
$$\text{By Cauchy-Schwarz: } E[XY] \leq \sqrt{E[X^2]E[Y^2]}$$
$$E[X^2]E[Y^2] \geq E[XY]^2 = E[X(1 - 1_{x\leq t})]^2$$
$$= (1 - E[X1_{X\leq t}])^2$$
$$\geq (1-t)^2$$

Definition 1.15.1: Paykey-Zygmund inequality

for $t \in (0,1)$

$$P(Y > tE[Y]) \geq (1-t)^2 \frac{(E[Y])^2}{E[Y^2]}$$

second moment method

§1.16 Lecture 12 (03-10)

Midterm March 26

Definition 1.16.1: Monetone Convergence Theorem

For any increasing sequence of function:$\{f_n\}$ such that $\{f_n\}$ is bounded from below
Then

$$\lim_{n\to+\infty} \int f_n d\mu = \int \lim_{n\to+\infty} f_n d\mu$$

: **Proof:**
Since $\int f_n d\mu \leq \int f\mu \forall n \in N$ Then $\limsup \int f_n d\mu \leq f d\mu$
Fact:
if ϕ is a simple function, then $u(A) := \int_A d\mu$ defines a measure(ex.)
Take a sequence $\{\phi_k\}$ of the simple function: $\phi_k \nearrow f$
Let $\alpha \in (0,1)$ Fix a given ϕ_k
Let $A_n = \{f_n > \alpha\phi_k\} = \{\omega \in \Omega : f_n(\omega) > \alpha\phi_k(\omega)\}$

Then $(A_n) \nearrow$, and

$$\begin{aligned}\lim_{n\to+\infty} &= \bigcup_{n\geq 1} A_n = \{\exists n \in N s.t. f_n > \alpha\phi_k\} \\ &= \{sup_n f_n > \alpha\phi k\} = \Omega \\ \text{Since } \int_\Omega f_n d\mu &\geq \int_{A_n} f_n d\mu \geq \alpha \int_{A_n} \phi_k d\mu \\ \text{send } n \to +\infty \liminf \int_\Omega f_n d\mu &\geq \alpha \lim_{n\to+\infty} \int_{A_n} \phi_k d\mu = \alpha \int_\Omega \phi_k d\mu \\ \text{Send } \phi_k \to f &\text{ and } \alpha \to 1 \text{ to complete the pf}\end{aligned}$$

Proposition 1.16.0

Let f, g be Lebesgue measurable. Then

- $$\int_\Omega af + bg d\mu = a\int_\Omega f d\mu + b\int_\Omega g d\mu$$
- if $f \leq g$ then
$$\int_\Omega f d\mu \leq \int_\Omega g d\mu$$

Proposition 1.16.0

f is (Lebesgue) integrable i |f| is integrable.
If there exists a function Y, s.t. $|f| \leq Y$, and Y is integrable, then f is integrable.

Proof:

$$\begin{aligned}&\text{Note that } f = f^+ - f^-,\ |f| = f^+ + f^- \\ &\text{Therefore } \int_\Omega |f| d\mu < \infty \Leftrightarrow \int_\Omega f^+ d\mu < \infty \text{ and } \int_\Omega f^- d\mu < \infty \Leftrightarrow \int_\Omega f d\mu < \infty.\end{aligned}$$

$$\text{If } |f| \leq Y, \text{ then } \int_\Omega |f| d\mu \leq \int_\Omega Y d\mu < \infty \Rightarrow \text{ |f| integrable } \Rightarrow f \text{ integrable.}$$

eg. $(R, B(R), m)$, $f(x) = \sin x$. Is f Lebesgue integrable?

$$|f(x)| \leq 1 \Rightarrow \int_R |f| dx \geq \sum_k \int_{k\pi+\frac{\pi}{6}}^{(k+1)\pi-\frac{\pi}{6}} |sinx| dx = +\infty$$

$\Rightarrow f$ is NOT integrable

Definition 1.16.2

We say $f = g$ almost everywhere (a.e.) if $\{f \neq g\}$ has measure 0.
r.v.s. $X = Y$ almost surely (a.s.) if $\mathbb{P}(\{X \neq Y\}) = 0$

Theorem 1.16.1

f $\mu(A) = 0$, then $\int_A f\, d\mu = 0$ for any measurable function f.

Proof. We prove this in three steps:

Step 1: Simple functions.

For a simple function $s = \sum_{i=1}^n a_i \chi_{E_i}$, where χ_{E_i} is the characteristic function of set E_i:

$$\int_A s\, d\mu = \sum_{i=1}^n a_i \mu(A \cap E_i)$$

Since $\mu(A) = 0$, we have $\mu(A \cap E_i) \leq \mu(A) = 0$ for any measurable set E_i. Therefore, $\mu(A \cap E_i) = 0$.

Hence, $\int_A s\, d\mu = \sum_{i=1}^n a_i \cdot 0 = 0$

Step 2: Non-negative measurable functions.

For any non-negative measurable function $f \geq 0$, there exists an increasing sequence of simple functions $\{s_n\}$ such that $s_n \uparrow f$ pointwise.

By the Monotone Convergence Theorem:

$$\int_A f\, d\mu = \lim_{n\to\infty} \int_A s_n\, d\mu$$

From Step 1, for each n, $\int_A s_n\, d\mu = 0$. Therefore, $\int_A f\, d\mu = \lim_{n\to\infty} 0 = 0$

Step 3: General measurable functions.

For any measurable function f, we can decompose it as $f = f^+ - f^-$, where $f^+ = \max(f, 0)$ and $f^- = \max(-f, 0)$ are both non-negative measurable functions.

By the linearity of the integral:

$$\int_A f\, d\mu = \int_A f^+\, d\mu - \int_A f^-\, d\mu$$

From Step 2, $\int_A f^+\, d\mu = 0$ and $\int_A f^-\, d\mu = 0$. Therefore, $\int_A f\, d\mu = 0 - 0 = 0$

Thus, if $\mu(A) = 0$, then $\int_A f\, d\mu = 0$ for any measurable function f. ∎

Corollary 1.16.1

f $f = g$ almost everywhere (a.e.), then $\int_\Omega f\, d\mu = \int_\Omega g\, d\mu$.

Proof. Let $E = \{x \in \Omega : f(x) \neq g(x)\}$. Since $f = g$ a.e., we have $\mu(E) = 0$.

Consider $h = f - g$. Then $h = 0$ on $\Omega \setminus E$, and $h \neq 0$ only on E.

Therefore:

$$\int_\Omega (f - g)\, d\mu = \int_\Omega h\, d\mu = \int_E h\, d\mu + \int_{\Omega\setminus E} h\, d\mu = \int_E h\, d\mu + 0$$

Since $\mu(E) = 0$, by our theorem, $\int_E h\, d\mu = 0$. Thus, $\int_\Omega (f - g)\, d\mu = 0$.

By the linearity of the integral:

$$\int_\Omega f\,d\mu - \int_\Omega g\,d\mu = \int_\Omega (f-g)\,d\mu = 0$$

Therefore, $\int_\Omega f\,d\mu = \int_\Omega g\,d\mu$. ■

Proposition 1.16.0

Let $(\Omega, \mathcal{F}, \mu)$ be a measure space. Let $f : \Omega \to [0, +\infty)$ be a Borel function.

Then $\nu(A) = \int_A f\,d\mu$, $\forall A \in \mathcal{F}$, defines a measure.

Definition 1.16.3

We say f is the Radon-Nikodym derivative (or density) of ν with respect to μ. Write $f = \frac{d\nu}{d\mu}$.

Proof of Proposition.

- $\nu(\emptyset) = 0$ is obvious.
- Countable additivity: Let $(A_i)_{i=1}^{+\infty}$ be disjoint. Let $A = \cup_{i=1}^{+\infty} A_i$.

  Then:

$$\begin{aligned}
\nu(A) &= \int_A f\,d\mu \\
&= \int_\Omega f \cdot \mathbf{1}_A\,d\mu \\
&= \int_\Omega f \cdot \left(\lim_{n\to\infty} \sum_{i=1}^n \mathbf{1}_{A_i}\right) d\mu \\
&\overset{\text{MCT}}{=} \lim_{n\to\infty} \int_\Omega f \cdot \left(\sum_{i=1}^n \mathbf{1}_{A_i}\right) d\mu \\
&= \lim_{n\to\infty} \sum_{i=1}^n \int_\Omega f \cdot \mathbf{1}_{A_i}\,d\mu \\
&= \lim_{n\to\infty} \sum_{i=1}^n \int_{A_i} f\,d\mu \\
&= \lim_{n\to\infty} \sum_{i=1}^n \nu(A_i) \\
&= \sum_{i=1}^{+\infty} \nu(A_i)
\end{aligned}$$

  Therefore, ν is a measure.

■

Definition 1.16.4

We say ν is absolutely continuous with respect to μ (denoted $\nu \ll \mu$) if for any $A \in \mathcal{F}$ such that $\mu(A) = 0$, then $\nu(A) = 0$.

Example 1

If $\nu(A) = \int_A f \, d\mu$ for some non-negative Borel function f, then $\mu(A) = 0 \Rightarrow \nu(A) = \int_A f \, d\mu = 0$. Therefore, $\nu \ll \mu$.

Example 2: Lebesgue Measure Equivalence

$M_{Leb} \ll 2M_{Leb}$ and $2M_{Leb} \ll M_{Leb}$

Definition 1.16.5: Lebesgue Measure

The Lebesgue measure, denoted by m or λ, is a complete measure on the σ-algebra of Lebesgue measurable subsets of $\mathbb{R}^n$ that satisfies:

1. (Normalization) The measure of the unit cube is 1: $m([0,1]^n) = 1$.
2. (Translation invariance) For any measurable set E and any point $x \in \mathbb{R}^n$, $m(E + x) = m(E)$, where $E + x = \{y + x : y \in E\}$.
3. (Countable additivity) For any countable collection $\{E_i\}_{i=1}^{\infty}$ of pairwise disjoint measurable sets, $m(\cup_{i=1}^{\infty} E_i) = \sum_{i=1}^{\infty} m(E_i)$.

Theorem 1.16.2: Properties of Lebesgue Measure

The Lebesgue measure has the following properties:

1. For an interval $[a, b] \subset \mathbb{R}$, $m([a, b]) = b - a$.
2. More generally, for a rectangle $[a_1, b_1] \times \cdots \times [a_n, b_n] \subset \mathbb{R}^n$, $m([a_1, b_1] \times \cdots \times [a_n, b_n]) = \prod_{i=1}^{n} (b_i - a_i)$.
3. For any $c \in \mathbb{R}$ and any measurable set $E \subset \mathbb{R}^n$, $m(cE) = |c|^n m(E)$, where $cE = \{cx : x \in E\}$.
4. There exist subsets of $\mathbb{R}$ that are not Lebesgue measurable.

Definition 1.16.6: Equivalent Measures

If $\mu \ll \nu$ and $\nu \ll \mu$, then we say μ and ν are equivalent (denoted $\mu \sim \nu$).

Example 3: Dirac Measures and Counting Measure

Consider $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$: Dirac measure. For each $k \in \mathbb{N}$, define μ_k such that

$$\mu_k(A) = \begin{cases} 1 & \text{if } k \in A \\ 0 & \text{else} \end{cases}$$

Counting measure: $\nu(A) = \sum_{k\in\mathbb{N}} \mu_k(A)$ (= "number of elements in A").
Therefore, $\mu_k \ll \nu$ for all $k \in \mathbb{N}$, but $\nu \not\ll \mu_k$.

Theorem 1.16.3: Radon-Nikodym Theorem

If μ, ν are σ-finite measures, and $\nu \ll \mu$, then there exists a Borel function f, such that

$$\forall A \in \mathcal{F}, \quad \nu(A) = \int_A f\, d\mu.$$

Proposition 1.16.0: Equivalent Characterization of Absolute Continuity

$\nu \ll \mu \Leftrightarrow \forall \varepsilon > 0, \exists \delta > 0,$ s.t. $\forall A \in \mathcal{F},$ with $\mu(A) < \delta$ we have $\nu(A) < \varepsilon$.

Proof. We will prove both directions of the equivalence.

($\Rightarrow$) **Necessity:** Suppose $\nu \ll \mu$. We need to show that for any $\varepsilon > 0$, there exists $\delta > 0$ such that for all $A \in \mathcal{F}$ with $\mu(A) < \delta$, we have $\nu(A) < \varepsilon$.

We will prove this by contradiction. Suppose, contrary to our claim, that there exists some $\varepsilon > 0$ such that for all $n \in \mathbb{N}$, there exists $A_n \in \mathcal{F}$ with $\mu(A_n) < \frac{1}{2^n}$ but $\nu(A_n) \geq \varepsilon$.

Let us define $B_k = \bigcup_{n=k}^{\infty} A_n$ for each $k \in \mathbb{N}$. Then $\{B_k\}_{k=1}^{\infty}$ forms a decreasing sequence of sets whose limit is $\lim_{k\to\infty} B_k = \bigcap_{k=1}^{\infty} B_k = \liminf_{n\to\infty} A_n$.

For each k, we can estimate:

$$\begin{aligned} \mu(B_k) &\leq \sum_{n=k}^{\infty} \mu(A_n) \\ &< \sum_{n=k}^{\infty} \frac{1}{2^n} \\ &= \frac{1}{2^{k-1}} \end{aligned}$$

Since $A_k \subseteq B_k$ for each k, and $\nu(A_k) \geq \varepsilon$, it follows that $\nu(B_k) \geq \varepsilon$ for all k.
Now, let $B = \bigcap_{k=1}^{\infty} B_k$. By the continuity of measure for decreasing sequences:

$$\begin{aligned} \mu(B) &= \lim_{k\to\infty} \mu(B_k) \\ &\leq \lim_{k\to\infty} \frac{1}{2^{k-1}} \\ &= 0 \end{aligned}$$

Similarly, by the continuity of measure:

$$\begin{aligned} \nu(B) &= \lim_{k\to\infty} \nu(B_k) \\ &\geq \varepsilon > 0 \end{aligned}$$

This gives us a set B with $\mu(B) = 0$ but $\nu(B) \geq \varepsilon > 0$, which contradicts our assumption that $\nu \ll \mu$. Therefore, our original claim must be true.

($\Leftarrow$) Sufficiency: Suppose that for any $\varepsilon > 0$, there exists $\delta > 0$ such that for all $A \in \mathcal{F}$ with $\mu(A) < \delta$, we have $\nu(A) < \varepsilon$.

We need to show that $\nu \ll \mu$, that is, for any $A \in \mathcal{F}$ with $\mu(A) = 0$, we have $\nu(A) = 0$.

Let $A \in \mathcal{F}$ with $\mu(A) = 0$. For any $\varepsilon > 0$, by our assumption, there exists a $\delta > 0$ such that for all $E \in \mathcal{F}$ with $\mu(E) < \delta$, we have $\nu(E) < \varepsilon$.

Since $\mu(A) = 0 < \delta$, it follows that $\nu(A) < \varepsilon$. But this is true for any $\varepsilon > 0$, no matter how small. Therefore, $\nu(A) = 0$, which proves that $\nu \ll \mu$. ■

§1.17 Lecture 13 (03-12)

Definition 1.17.1: Radon-Nikodym Derivative

If $\nu(A) = \int_A f \, d\mu$ for some nonnegative Borel measurable function f, then f is the Radon-Nikodym derivative of ν with respect to μ.

Definition 1.17.2: Absolute Continuity

$\nu \ll \mu$ if and only if $\forall A \in \mathcal{F}$ such that $\mu(A) = 0 \Rightarrow \nu(A) = 0$.

Theorem 1.17.1: Radon-Nikodym Theorem

If μ, ν are σ-finite measures, then $\nu \ll \mu \Leftrightarrow \nu(A) = \int_A f \, d\mu, \forall A \in \mathcal{F}$ for some nonnegative measurable function f.

Definition 1.17.3: Equivalent Measures

$\mu \sim \nu$ if and only if $\mu \ll \nu$ and $\nu \ll \mu$.

Proposition 1.17.0: Continuous Random Variables

A continuous random variable X satisfies $\mathbb{P}_X(A) = \mathbb{P}(X \in A), \forall A \in \mathcal{B}(\mathbb{R})$.

$\mathbb{P}_X \ll m \Leftrightarrow \exists f$ such that $\mathbb{P}_X(A) = \int_A f(x) \, dm$, where f is the probability density function.

Remark 1

We often write $\int_A f \, dm = \int_A f(x) \, dx$ when m is the Lebesgue measure.

Example 1: Exponential Distribution

Consider the measure space $(\mathbb{R}, \mathcal{B}(\mathbb{R}), m)$ where m is the Lebesgue measure. Define the function:

$$g(x) = \lambda e^{-\lambda x} \mathbf{1}_{[0,+\infty)}(x), \quad \lambda > 0$$

Then for every $B \in \mathcal{B}(\mathbb{R})$,

$$\mathbb{P}_X(B) = \int_B g(x)\, dx$$

defines a measure.

To verify this is a probability measure, we check:

$$\begin{aligned}
\mathbb{P}_X(\mathbb{R}) &= \int_{-\infty}^{+\infty} \lambda e^{-\lambda x} \mathbf{1}_{[0,+\infty)}(x)\, dx \\
&= \int_0^{+\infty} \lambda e^{-\lambda x}\, dx \\
&= \left[-e^{-\lambda x}\right]_0^{+\infty} \\
&= 0 - (-1) = 1
\end{aligned}$$

Therefore, $\mathbb{P}_X$ defines a probability measure.

This is the **exponential distribution** with parameter λ, denoted as $X \sim \mathrm{Exp}(\lambda)$.

Fact: If $X \sim \mathrm{Exp}(\lambda)$, then:

$$\begin{aligned}
\mathbb{E}[X] &= \int_0^{+\infty} x \lambda e^{-\lambda x}\, dx \\
&= -\int_0^{+\infty} x\, d(e^{-\lambda x}) \\
&= -\left[x e^{-\lambda x}\right]_0^{+\infty} + \int_0^{+\infty} e^{-\lambda x}\, dx \\
&= 0 - 0 + \frac{1}{\lambda}\left[-e^{-\lambda x}\right]_0^{+\infty} \\
&= \frac{1}{\lambda}
\end{aligned}$$

Also:

$$\begin{aligned}
\mathbb{E}[X^2] &= \int_0^{+\infty} x^2 \lambda e^{-\lambda x}\, dx \\
&= \dots \\
&= \frac{2}{\lambda^2}
\end{aligned}$$

Lemma 1.17.1: Standardization of Normal Distribution

If $X \sim N(\mu, \sigma^2)$, then $Y = \frac{X-\mu}{\sigma}$ is $N(0,1)$ (standard normal).

Proof.

$$\mathbb{P}(Y \leq a) = \mathbb{P}\left(X \leq \mu + a\sigma\right) = \int_{-\infty}^{\mu+a\sigma} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\, dx$$

■

Making the substitution $y = \frac{x-\mu}{\sigma}$, we get:

$$\mathbb{P}(Y \leq a) = \int_{-\infty}^{a} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}\, dy$$

which is the CDF of a standard normal distribution.

Lemma 1.17.2: Moment Generating Function of Standard Normal

If $X \sim N(0,1)$, then $\mathbb{E}[e^{tX}] = e^{\frac{1}{2}t^2}$.

Proof.

$$\begin{aligned}
\mathbb{E}[e^{tX}] &= \int_{-\infty}^{+\infty} e^{tx} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\, dx \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}(x^2-2tx)}\, dx \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{(x-t)^2}{2}} e^{\frac{t^2}{2}}\, dx \\
&= e^{\frac{t^2}{2}} \cdot \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{(x-t)^2}{2}}\, dx \\
&= e^{\frac{t^2}{2}} \cdot 1 \\
&= e^{\frac{t^2}{2}}
\end{aligned}$$

where in the fourth line we used the substitution $y = x - t$. ■

Remark 2: Applications of MGF

(1) If $\mathbb{E}[e^{tX}] < +\infty$, then:

$$\begin{aligned}
\mathbb{E}[e^{tX}] &= \mathbb{E}\left[\sum_{k=0}^{\infty} \frac{t^k}{k!} X^k\right] \\
&= \sum_{k=0}^{\infty} \frac{t^k}{k!} \mathbb{E}[X^k]
\end{aligned}$$

Therefore, the moment generating function determines all moments of the random variable.

(2) Let X be a continuous random variable with density function f. Then:

$$\mathbb{E}[e^{tX}] = \int_{-\infty}^{+\infty} e^{tx} f(x)\, dx = \mathcal{L}[f]$$

where $\mathcal{L}[f]$ denotes the Laplace transform of f.

(3) Fact: If $M_X(t) = \mathbb{E}[e^{tX}]$ converges absolutely in a neighborhood of 0, then it uniquely determines the probability density function f.
(4) If $X \sim N(\mu, \sigma^2)$, compute $\mathbb{E}[e^{tX}]$.

Proposition 1.17.0: Lack of memory

If $X \sim Exp(\lambda)$, Then

$$P(X > s + t | X > s) = P(X > t)$$

Proof.

$$\begin{aligned}
\mathbb{P}[X > t + s | X > s] &= \frac{\mathbb{P}[X > t + s, X > s]}{\mathbb{P}[X > s]} \\
&= \frac{\mathbb{P}[X > t + s]}{\mathbb{P}[X > s]} \\
&= \frac{\int_{t+s}^{+\infty} \lambda e^{-\lambda x}\, dx}{\int_{s}^{+\infty} \lambda e^{-\lambda x}\, dx} \\
&= \frac{e^{-\lambda(t+s)}}{e^{-\lambda s}} \\
&= e^{-\lambda t} \\
&= \mathbb{P}[X > t]
\end{aligned}$$

■

Example 2: Probability Density Function of Normal Distribution

The probability density function of a normal distribution is given by:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \mu \in \mathbb{R}, \sigma > 0$$

This is also known as the Gaussian distribution, denoted as $N(\mu, \sigma^2)$.
Special case: When $\mu = 0$ and $\sigma = 1$, we have:

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

which is called the standard normal distribution, denoted as $N(0, 1)$.

Proposition 1.17.0: Properties of Normal Distribution

1. **Normalization:** Verify that $\int_{\mathbb{R}} f(x)\, dx = 1$, since

$$\int_{-\infty}^{+\infty} e^{-\frac{x^2}{2}}\, dx = \sqrt{2\pi}$$

2. **Mean and Variance:** If $X \sim N(\mu, \sigma^2)$, then $\mathbb{E}[X] = \mu$ and $\mathrm{Var}(X) = \sigma^2$.

3. **Special Case - Standard Normal Distribution:** If $X \sim N(0,1)$, then:

$$\begin{aligned}
\mathbb{E}[X] &= \int_{-\infty}^{+\infty} x \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\, dx = 0 \\
\mathrm{Var}(X) = \mathbb{E}[X^2] &= \int_{-\infty}^{+\infty} x^2 \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\, dx \\
&= \frac{1}{\sqrt{2\pi}} \frac{d}{d\lambda} \left[\int_{-\infty}^{+\infty} e^{-\lambda x^2}\, dx \right]_{\lambda=\frac{1}{2}} \\
&= \frac{1}{\sqrt{2\pi}} \frac{d}{d\lambda} \left[\sqrt{\frac{\pi}{\lambda}} \right]_{\lambda=\frac{1}{2}} \\
&= \frac{1}{\sqrt{2\pi}} \cdot \left(-\frac{1}{2}\right) \cdot \frac{d}{d\lambda} \left[\lambda^{-\frac{1}{2}}\right]_{\lambda=\frac{1}{2}} \cdot \sqrt{\pi} \\
&= \frac{1}{\sqrt{2\pi}} \cdot \left(-\frac{1}{2}\right) \cdot \left(-\frac{1}{2}\right) \cdot \lambda^{-\frac{3}{2}} \cdot \sqrt{\pi}\Big|_{\lambda=\frac{1}{2}} \\
&= \frac{1}{4} \cdot \frac{\sqrt{2\pi}}{\sqrt{2\pi}} \cdot 2\sqrt{2} \\
&= 1
\end{aligned}$$

Example 3: Cauchy Distribution

Let $f(x) = \frac{1}{\pi(1+x^2)}, \quad x \in \mathbb{R}$.

$$\begin{aligned}
\int_{-\infty}^{+\infty} f(x)\, dx &= \int_{-\infty}^{+\infty} \frac{1}{\pi(1+x^2)}\, dx \\
&= \frac{1}{\pi} [\arctan(x)]_{-\infty}^{+\infty} \\
&= \frac{1}{\pi} \cdot \pi = 1
\end{aligned}$$

This defines a random variable $Y \sim \mathrm{Cauchy}(0,1)$.
Fact: $\mathbb{E}[e^{tX}] = \int \frac{e^{tx}}{\pi(1+x^2)}\, dx$ converges only for $t = 0$.
The expected value is also divergent:

$$\mathbb{E}[X] = \int \frac{x}{\pi(1+x^2)}\, dx \quad \text{diverges}$$

Characteristic Function: While the moment generating function doesn't exist for the Cauchy distribution, the characteristic function is well-defined:

$$\phi(t) = \mathbb{E}[e^{itX}] = \int e^{itx} \frac{1}{\pi(1+x^2)}\, dx$$

§1.18 Recitation (03-14)

Problem 1

We know that $X \sim Exp(\lambda)$,Then $P(X > t + s|X > s) = P(X > t)$
Identify all distribution that satisfy the lack of memory property.

Proof. We proceed in several steps to identify all distributions that satisfy the memoryless property.

Step 1: First, we recall the definition of the memoryless property. A random variable X has the memoryless property if

$$\mathbb{P}(X > t + s|X > s) = \mathbb{P}(X > t) \quad \forall s, t \geq 0 \tag{1}$$

We know that if $X \sim \text{Exp}(\lambda)$ (i.e., X follows an exponential distribution with parameter λ), then X satisfies property (1). We want to prove that exponential distributions are the only ones with this property.

Step 2: Let $G(t) = \mathbb{P}(X > t)$ be the survival function of X. Using property (1), we have:

$$\begin{aligned} \mathbb{P}(X > t + s|X > s) &= \mathbb{P}(X > t) \\ \frac{\mathbb{P}(X > t + s, X > s)}{\mathbb{P}(X > s)} &= \mathbb{P}(X > t) \\ \frac{\mathbb{P}(X > t + s)}{\mathbb{P}(X > s)} &= \mathbb{P}(X > t) \\ \frac{G(t + s)}{G(s)} &= G(t) \end{aligned}$$

Rearranging, we get the functional equation:

$$G(t + s) = G(t) \cdot G(s) \quad \forall s, t \geq 0 \tag{2}$$

Step 3: Define $H(t) = -\log G(t)$. Taking the logarithm of both sides of equation (2):

$$\log G(t + s) = \log G(t) + \log G(s) \tag{1.4}$$
$$-H(t + s) = -H(t) - H(s) \tag{1.5}$$
$$H(t + s) = H(t) + H(s) \quad \forall s, t \geq 0 \tag{3}$$

This is the Cauchy functional equation. We will now prove that any continuous solution to this equation must have the form $H(t) = \lambda t$ for some constant $\lambda > 0$.

Step 4: We claim that there exists λ such that $H(t) = \lambda t$ for all $t \geq 0$. We prove this in several sub-steps:

(a) For integer values: Let $\lambda = H(1)$. For any integer $k \in \mathbb{Z}^+$, we have:

$$\begin{aligned} H(k) &= H(1 + 1 + \ldots + 1) \quad (k \text{ times}) \\ &= H(1) + H(1) + \ldots + H(1) \quad (\text{using equation (3) repeatedly}) \\ &= k \cdot H(1) = \lambda k \end{aligned}$$

(b) For rational values: For any rational number $\frac{p}{q}$ where $p, q \in \mathbb{Z}^+$, using equation

(3), we have:

$$H\left(\frac{p}{q} \cdot q\right) = q \cdot H\left(\frac{p}{q}\right)$$
$$H(p) = q \cdot H\left(\frac{p}{q}\right)$$
$$\lambda p = q \cdot H\left(\frac{p}{q}\right)$$
$$H\left(\frac{p}{q}\right) = \lambda \cdot \frac{p}{q}$$

(c) For all real values: By the continuity of H and the density of rational numbers in $\mathbb{R}$, for any real number $t \geq 0$, there exists a sequence of rational numbers $\{t_n\}$ such that $t_n \to t$. By the continuity of H, we have:

$$H(t) = \lim_{n\to\infty} H(t_n) = \lim_{n\to\infty} \lambda t_n = \lambda t$$

Therefore, $H(t) = \lambda t$ for all $t \geq 0$.
Step 5: Since $H(t) = -\log G(t) = \lambda t$, we have:

$$G(t) = e^{-\lambda t}, \quad \lambda > 0$$

This is exactly the survival function of an exponential distribution with parameter λ. Note that λ must be positive since $G(t)$ is a decreasing function of t (as t increases, the probability of surviving beyond t decreases).

Step 6: Finally, the probability density function of X is the negative derivative of the survival function:

$$f(t) = -\frac{d}{dt}G(t) = -\frac{d}{dt}e^{-\lambda t} = \lambda e^{-\lambda t}, \quad t \geq 0$$

This is the probability density function of an exponential distribution with parameter λ.

In conclusion, the only continuous probability distributions that satisfy the memoryless property are exponential distributions. ■

Problem 2

(Ω, F, μ) measure space, X is a integrable borel function, Y is an simple function
Show that (1)

$$\exists A \in F, \mu(A) < +\infty, s.t. \int_\Omega Y d\mu = \int_A Y d\mu$$

(2)

$$\exists \epsilon > 0, \exists A_\epsilon \in F, \mu(A_\epsilon) < +\infty, s.t. |\int_{A_\epsilon} X d\mu - \int_\Omega X d\mu| < \epsilon$$

Proof. **Part 1:** Let Y be a simple function, which can be written as

$$Y = \sum_{i=1}^{n} a_i 1_{A_i}$$

where $a_i \in \mathbb{R}$, $A_i \cap A_j = \emptyset$ for $i \neq j$, and $a_i \neq 0$. Let $A = \bigcup_{i=1}^{n} A_i$.

Since Y is zero outside A, we have

$$\int_\Omega Y\,d\mu = \int_A Y\,d\mu$$

Also, $\mu(A) = \sum_{i=1}^n \mu(A_i) < +\infty$ since each A_i has finite measure.

Part 2: Suppose that $X \geq 0$ is a non-negative Borel function. Since X is integrable, there exists a sequence of simple functions $\{Y_n\}$ such that $Y_n \uparrow X$ pointwise and $\lim_{n\to\infty} \int_\Omega Y_n\,d\mu = \int_\Omega X\,d\mu$.

For any $\epsilon > 0$, there exists $N \in \mathbb{N}$ such that

$$\left|\int_\Omega Y_N\,d\mu - \int_\Omega X\,d\mu\right| < \epsilon$$

From Part 1, there exists $A_N \in \mathcal{F}$ with $\mu(A_N) < +\infty$ such that

$$\int_\Omega Y_N\,d\mu = \int_{A_N} Y_N\,d\mu$$

This implies

$$\left|\int_{A_N} Y_N\,d\mu - \int_\Omega X\,d\mu\right| < \epsilon$$

Since $0 \leq Y_N \leq X$ and Y_N approximates X on A_N, we have

$$\int_{A_N} X\,d\mu - \epsilon \leq \int_{A_N} Y_N\,d\mu \leq \int_{A_N} X\,d\mu$$

Which leads to

$$\left|\int_{A_N} X\,d\mu - \int_\Omega X\,d\mu\right| < \epsilon$$

Now, suppose X is an arbitrary integrable Borel function. Let $X = X^+ - X^-$, where $X^+ = \max(X, 0)$ and $X^- = \max(-X, 0)$ are the positive and negative parts of X, respectively.

By applying the above result to X^+ and X^-, we can find $A^+, A^- \in \mathcal{F}$ with $\mu(A^+), \mu(A^-) < +\infty$ such that

$$\left|\int_{A^+} X^+\,d\mu - \int_\Omega X^+\,d\mu\right| < \frac{\epsilon}{2}$$
$$\left|\int_{A^-} X^-\,d\mu - \int_\Omega X^-\,d\mu\right| < \frac{\epsilon}{2}$$

Let $A = A^+ \cup A^-$. Then $\mu(A) < +\infty$ and

$$\begin{aligned}
\left|\int_A X\,d\mu - \int_\Omega X\,d\mu\right| &= \left|\int_A (X^+ - X^-)\,d\mu - \int_\Omega (X^+ - X^-)\,d\mu\right| \\
&\leq \left|\int_A X^+\,d\mu - \int_\Omega X^+\,d\mu\right| + \left|\int_A X^-\,d\mu - \int_\Omega X^-\,d\mu\right| \\
&< \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon
\end{aligned}$$

Therefore, for any integrable Borel function X and any $\epsilon > 0$, there exists $A_\epsilon \in \mathcal{F}$ with $\mu(A_\epsilon) < +\infty$ such that

$$\left|\int_{A_\epsilon} X\,d\mu - \int_\Omega X\,d\mu\right| < \epsilon$$

■

Problem 3

Let X be a continuous non-negative random variable with $\mathbb{E}[X] < \infty$. Then:

1. $\mathbb{E}[X] = \int_0^{+\infty} \mathbb{P}[X > y]\, dy - \int_0^{+\infty} \mathbb{P}[X < -y]\, dy$
2. $\lim_{y \to +\infty} y \cdot \mathbb{P}[X > y] = 0$

Proof. Let's prove each part separately.

Part (1): We begin with the definition of expectation. For a continuous random variable X with probability density function $f_X(x)$:

$$\mathbb{E}[X] = \int_{-\infty}^{+\infty} x \cdot f_X(x)\, dx$$

We can decompose X into its positive and negative parts:

$$\mathbb{E}[X] = \mathbb{E}[X^+] - \mathbb{E}[X^-]$$

where $X^+ = \max(X, 0)$ and $X^- = \max(-X, 0)$.

In class, we showed that for a non-negative random variable Z:

$$\mathbb{E}[Z] = \int_0^{+\infty} \mathbb{P}[Z > y]\, dy$$

Applying this to X^+ and X^-:

$$\mathbb{E}[X^+] = \int_0^{+\infty} \mathbb{P}[X^+ > y]\, dy = \int_0^{+\infty} \mathbb{P}[X > y]\, dy$$
$$\mathbb{E}[X^-] = \int_0^{+\infty} \mathbb{P}[X^- > y]\, dy = \int_0^{+\infty} \mathbb{P}[-X > y]\, dy = \int_0^{+\infty} \mathbb{P}[X < -y]\, dy$$

Therefore:

$$\begin{aligned}\mathbb{E}[X] &= \mathbb{E}[X^+] - \mathbb{E}[X^-] \\ &= \int_0^{+\infty} \mathbb{P}[X > y]\, dy - \int_0^{+\infty} \mathbb{P}[X < -y]\, dy\end{aligned}$$

Part (2): We need to prove that $\lim_{y \to +\infty} y \cdot \mathbb{P}[X > y] = 0$.

For any $y > 0$, we have:

$$\begin{aligned}y \cdot \mathbb{P}[X > y] &= y \cdot \int_y^{+\infty} f_X(x)\, dx \\ &\leq \int_y^{+\infty} x \cdot f_X(x)\, dx \\ &= \mathbb{E}[X \cdot \mathbf{1}_{\{X > y\}}]\end{aligned}$$

We claim that $\mathbb{E}[X \cdot \mathbf{1}_{\{X > y\}}] \to 0$ as $y \to +\infty$.

This follows because X is integrable ($\mathbb{E}[X] < \infty$), and as y increases, the set $\{X > y\}$ becomes smaller. By the dominated convergence theorem:

$$\lim_{y \to +\infty} \mathbb{E}[X \cdot \mathbf{1}_{\{X > y\}}] = 0$$

Therefore:

$$\lim_{y\to+\infty} y \cdot \mathbb{P}[X > y] = 0$$

This result can also be approached using approximation by simple functions. Additionally, we can leverage the following fact from measure theory: In a measure space $(\Omega, \mathcal{F}, \mu)$, if f is integrable, then for any $\varepsilon > 0$, there exists $\delta > 0$ such that for any $A \in \mathcal{F}$ with $\mu(A) < \delta$, we have $\left|\int_A f\, d\mu\right| < \varepsilon$. ■

§1.19 Lecture 14 (03-17)

Theorem 1.19.1: Chain Rule for Radon-Nikodym Derivatives

Suppose $\nu \ll \mu$ (i.e., ν is absolutely continuous with respect to μ). Then a function f is integrable with respect to ν if and only if $f \cdot \frac{d\nu}{d\mu}$ is integrable with respect to μ. Furthermore, in this case:

$$\int f\, d\nu = \int f \cdot \frac{d\nu}{d\mu}\, d\mu$$

where $\frac{d\nu}{d\mu}$ denotes the Radon-Nikodym derivative of ν with respect to μ.

Remark 1: Probabilistic Interpretation

If $Q \ll P$ are probability measures, and X is a Q-integrable random variable, then:

$$\mathbb{E}^Q[X] = \mathbb{E}^P\left[X \cdot \frac{dQ}{dP}\right]$$

where $\frac{dQ}{dP}$ is the Radon-Nikodym derivative (likelihood ratio) of Q with respect to P.

Proof. We proceed in several steps:

Step 1: First, we verify the theorem for indicator functions. For any measurable set A:

$$\begin{aligned}\int 1_A\, d\nu &= \nu(A) \\ &= \int_A \frac{d\nu}{d\mu}\, d\mu \quad \text{(by definition of Radon-Nikodym derivative)} \\ &= \int 1_A \cdot \frac{d\nu}{d\mu}\, d\mu\end{aligned}$$

Step 2: By linearity, we extend the result to simple functions. Let $\varphi = \sum_{i=1}^n a_i 1_{A_i}$, where $A_i \cap A_j = \emptyset$ for $i \neq j$. Then:

$$\begin{aligned}
\int \varphi \, d\nu &= \sum_{i=1}^{n} a_i \int 1_{A_i} \, d\nu \\
&= \sum_{i=1}^{n} a_i \int 1_{A_i} \cdot \frac{d\nu}{d\mu} \, d\mu \quad \text{(by Step 1)} \\
&= \int \sum_{i=1}^{n} a_i 1_{A_i} \cdot \frac{d\nu}{d\mu} \, d\mu \\
&= \int \varphi \cdot \frac{d\nu}{d\mu} \, d\mu
\end{aligned}$$

Step 3: For a non-negative Borel function $g \geq 0$, there exists a sequence $\{\varphi_n\}_{n \geq 1}$ of simple functions such that $\varphi_n \uparrow g$ pointwise. Applying the Monotone Convergence Theorem (MCT):

$$\begin{aligned}
\int g \, d\nu &= \lim_{n \to \infty} \int \varphi_n \, d\nu \quad \text{(by MCT)} \\
&= \lim_{n \to \infty} \int \varphi_n \cdot \frac{d\nu}{d\mu} \, d\mu \quad \text{(by Step 2)} \\
&= \int \lim_{n \to \infty} \left(\varphi_n \cdot \frac{d\nu}{d\mu} \right) d\mu \quad \text{(by MCT)} \\
&= \int g \cdot \frac{d\nu}{d\mu} \, d\mu
\end{aligned}$$

Step 4: For a general Borel function g, we decompose it as $g = g^+ - g^-$, where $g^+ = \max(g, 0)$ and $g^- = \max(-g, 0)$. Applying the result from Step 3 to both g^+ and g^-, we obtain:

$$\begin{aligned}
\int g \, d\nu &= \int g^+ \, d\nu - \int g^- \, d\nu \\
&= \int g^+ \cdot \frac{d\nu}{d\mu} \, d\mu - \int g^- \cdot \frac{d\nu}{d\mu} \, d\mu \\
&= \int (g^+ - g^-) \cdot \frac{d\nu}{d\mu} \, d\mu \\
&= \int g \cdot \frac{d\nu}{d\mu} \, d\mu
\end{aligned}$$

This completes the proof of the chain rule. ■

Example 1

eg : $(\Omega, \mathcal{F}, \mathbb{P}), \quad X \sim N(0,1)$. Let $\theta > 0$. $\quad X + \theta \sim N(\theta, 1)$

We can define a new prob. measure $\mathbb{Q}$, s.t. under $\mathbb{Q}$, $X + \theta \sim N(0,1)$.

Soln : Let $\dfrac{d\mathbb{Q}}{d\mathbb{P}} = e^{-\theta X - \frac{1}{2}\theta^2}$

$\mathbb{Q}$ is a prob. measure: $\mathbb{Q}(\Omega) = \int_\Omega d\mathbb{Q} = \int_\Omega \frac{d\mathbb{Q}}{d\mathbb{P}} d\mathbb{P} = \int_\Omega e^{-\theta X - \frac{1}{2}\theta^2} d\mathbb{P} = e^{-\frac{1}{2}\theta^2}\mathbb{E}^{\mathbb{P}}[e^{-\theta X}]$

$$= e^{-\frac{1}{2}\theta^2} \cdot e^{\frac{1}{2}\theta^2} = 1 \quad \text{MGF for } N(0,1)$$

Compute $\mathbb{E}^{\mathbb{Q}}[e^{t(X+\theta)}] = \mathbb{E}^{\mathbb{P}}\left[e^{t(X+\theta)}\frac{d\mathbb{Q}}{d\mathbb{P}}\right]$

$\forall t \in \mathbb{R}$ $\qquad = \mathbb{E}^{\mathbb{P}}\left[e^{t(X+\theta)}e^{-\theta X - \frac{1}{2}\theta^2}\right] = e^{t\theta - \frac{1}{2}\theta^2}\mathbb{E}^{\mathbb{P}}[e^{(t-\theta)X}]$

$$= e^{t\theta - \frac{1}{2}\theta^2} e^{\frac{1}{2}(t-\theta)^2} = e^{\frac{1}{2}t^2}$$

$\Rightarrow X + \theta \sim^{\mathbb{Q}} N(0,1)$

Fact:

if $M(t) := \mathbb{E}[e^{tX}]$ converges in $t \in (-\delta, \delta)$ for some $\delta > 0$.
Then $\{M(t), t \in (-\delta, \delta)\}$ determines the distribution of X.

§1.19.1 Joint distribution

Given probability space $(\Omega_1, F_1, P_1), (\Omega_2, F_2, P_2)$ can define the product space $(\Omega_1 \times \Omega_2, F_1 \otimes F_2, P_1 \otimes P_2)$ where

$$F_1 \otimes F_2 = \sigma(\{A_1 \times A_2 : A_1 \in F_1, A_2 \in F_2\})$$

define

$$P_1 \otimes P_2(A_1 \times A_2) = P_1(A_1)P_2(A_2)$$

and extend to $F_1 \otimes F_2$ by Caratheodory extension theorem

Definition 1.19.1: Joint distribution

Joint distribution of X, Y is $F : \mathbb{R}^2 \to [0,1]$ s.t.

$$F(x,y) = P(X \le x, Y \le y)$$

Definition 1.19.2

If X,Y are continuous random variables, then the joint density function $f : \mathbb{R}^2 \to [0, +\infty)$ is given by

$$F(x,y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f(u,v)dudv$$

$$f(x,y) = \frac{\partial^2 F}{\partial x \partial y}$$

Remark 2

(1)

$$P[a < X \le b, c < Y \le d] = \int_a^b \int_c^d f(u,v)dudv$$

(2):by uniqueness of extension,

$$B(R^2) = \sigma\{\text{left open right closed cubes in } R^2\}$$
$$P((X,Y) \in A) = \int_A f(u,v)dudv, \text{ for every } A \in B(R^2)$$

May cover individual distribution from the joint distribution

$$F_X(x) = P(X \le x) = P(X \le x, Y \in R) = \int_{-\infty}^{x} \int_{-\infty}^{+\infty} f(u,v)dudv$$
$$f_X(x) = \int_{-\infty}^{+\infty} f(x,v)dv, f_Y(y) = \int_{-\infty}^{+\infty} f(u,y)du$$

Remark 3

if X,Y are independent$\Rightarrow$ $F(x,y) = F_X(x)F_Y(y) \Leftrightarrow (cont.) f(x,y) = f_X(x)f_Y(y)$

Example 2

①

If X, Y have joint density function

$f(x,y) = \frac{\alpha^x}{x!}\frac{\beta^y}{y!}e^{-\alpha-\beta} \quad \forall x, y \in \mathbb{N}.$

Soln: • X, Y are indep, because $f(x,y) = f_1(x)f_2(y)$.

• $f_X(x) = \sum_{y\in\mathbb{N}} \frac{\alpha^x}{x!}\frac{\beta^y}{y!}e^{-\alpha-\beta} = \frac{\alpha^x}{x!}e^{-\alpha-\beta}\sum_{y\in\mathbb{N}}\frac{\beta^y}{y!} = \frac{\alpha^x}{x!}e^{-\alpha}$

$\Rightarrow X \sim$ Poisson(α)

Similarly, $f_Y(y) = \sum_{x\in\mathbb{N}} \frac{\alpha^x}{x!}\frac{\beta^y}{y!}e^{-\alpha-\beta} = \dots \quad Y \sim$ Poisson(β).

②

The joint density function of X, Y is $f(x,y) = \begin{cases} e^{-(x+y)} & \text{if } 0 \le x, y < \infty \\ 0 & \text{else} \end{cases}$

Find the density funct. of $\frac{X}{Y}$.

Soln : $F_{\frac{X}{Y}}(a) = P\left(\frac{X}{Y} \le a\right) = \iint_{[x \le ay]} f(x,y)\,dx\,dy$

$= \int_0^{+\infty} \int_0^{ay} e^{-(x+y)}\,dx\,dy = \int_0^{+\infty} [1 - e^{-ay}]e^{-y}\,dy$

$= \int_0^{+\infty} (1 - e^{-ay})e^{-y}\,dy = 1 - \frac{1}{a+1}.$

$f_{\frac{X}{Y}}(a) = \frac{1}{(a+1)^2}, \quad a \ge 0$

§1.20 Lecture 15 (03-19)

§1.20.1 Joint Distribution Function and Density Function

Definition 1.20.1: Joint Distribution Function

For random variables X and Y, the joint distribution function is defined as:

$$F(x,y) = \mathbb{P}(X \le x, Y \le y)$$

Definition 1.20.2: Joint Density Function

For continuous random variables X and Y, if the second-order partial derivative of their joint distribution function $F(x,y)$ exists, the joint density function is defined as:

$$f(x,y) = \frac{\partial^2 F(x,y)}{\partial x \partial y}$$

For any measurable set $A \subset \mathbb{R}^2$, we have:

$$\mathbb{P}((X,Y) \in A) = \iint_A f(x,y)\,dx\,dy$$

§1.20.2 Distribution of Sum of Random Variables

Proposition 1.20.0: Density Function of Sum

Let X and Y be continuous random variables with joint density function $f(x,y)$.

The density function of $X+Y$ is:

$$f_{X+Y}(z) = \int_{-\infty}^{+\infty} f(x, z-x)\, dx$$

. First, the distribution function of $X+Y$ can be represented as:

$$\begin{aligned} F_{X+Y}(z) &= \mathbb{P}(X+Y \le z) \\ &= \iint_{x+y \le z} f(x,y)\, dx\, dy \\ &= \int_{-\infty}^{+\infty} \int_{-\infty}^{z-x} f(x,y)\, dy\, dx \end{aligned}$$

Differentiating with respect to z, we obtain the density function:

$$\begin{aligned} f_{X+Y}(z) &= \frac{d}{dz} F_{X+Y}(z) \\ &= \frac{d}{dz} \int_{-\infty}^{+\infty} \int_{-\infty}^{z-x} f(x,y)\, dy\, dx \\ &= \int_{-\infty}^{+\infty} f(x, z-x)\, dx \end{aligned}$$

where we applied Leibniz's integral rule to exchange the order of integration and differentiation. ■

Example 1: Sum of Independent Standard Normal Random Variables

Let $X, Y \sim N(0,1)$ be independent. Prove that $X+Y \sim N(0,2)$.

. Since X and Y are independent, their joint density function is the product of their individual density functions:

$$f(x,y) = f_X(x) \cdot f_Y(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}$$

Computing the density function of $X+Y$:

$$\begin{aligned} f_{X+Y}(z) &= \int_{-\infty}^{+\infty} f(x, z-x)\, dx \\ &= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{(z-x)^2}{2}}\, dx \\ &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}(x^2+(z-x)^2)}\, dx \end{aligned}$$

Computing the exponent term:

$$\begin{aligned} x^2 + (z-x)^2 &= x^2 + z^2 - 2zx + x^2 \\ &= 2x^2 - 2zx + z^2 \\ &= 2(x^2 - zx + \frac{z^2}{2}) \\ &= 2(x - \frac{z}{2})^2 + \frac{z^2}{2} - \frac{z^2}{2} \\ &= 2(x - \frac{z}{2})^2 + \frac{z^2}{2} \end{aligned}$$

Substituting back into the original integral:

$$f_{X+Y}(z) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-\frac{1}{2}[2(x-\frac{z}{2})^2+\frac{z^2}{2}]}\,dx$$
$$= \frac{1}{2\pi}e^{-\frac{z^2}{4}}\int_{-\infty}^{+\infty} e^{-(x-\frac{z}{2})^2}\,dx$$

Letting $u = x - \frac{z}{2}$, we have $dx = du$, and the integration limits transform:

$$f_{X+Y}(z) = \frac{1}{2\pi}e^{-\frac{z^2}{4}}\int_{-\infty}^{+\infty} e^{-u^2}\,du$$
$$= \frac{1}{2\pi}e^{-\frac{z^2}{4}}\cdot\sqrt{\pi}$$
$$= \frac{1}{\sqrt{2\pi\cdot 2}}e^{-\frac{z^2}{4}}$$

This is precisely the density function of $N(0,2)$, therefore $X+Y \sim N(0,2)$. ■

Fact:Sum of Independent Normal Random Variables
If $X \sim N(\mu_1,\sigma_1^2)$, $Y \sim N(\mu_2,\sigma_2^2)$ and X and Y are independent, then $X+Y \sim N(\mu_1+\mu_2,\sigma_1^2+\sigma_2^2)$.

Soln2: Use that moment generating funct. $M(t) = \mathbb{E}[e^{tX}]$ determines the distr.

$$M_{X+Y}(t) = \mathbb{E}[e^{t(X+Y)}] \overset{\text{ind.}}{=} \mathbb{E}[e^{tX}\cdot\mathbb{E}[e^{tY}]] = e^{\frac{1}{2}t^2}\cdot e^{\frac{1}{2}t^2} = e^{t^2}, \quad X+Y\sim N(0,2)$$

Here we used if $X \sim N(\mu,\sigma^2), \mathbb{E}[e^{tX}] = e^{\mu t+\frac{1}{2}\sigma^2 t^2}$.

Example 2: Problems with Uniform Distributions

Let X, Y be independent random variables, each uniformly distributed on $[0,1]$ (i.e., $X, Y \sim \text{Uniform}[0,1]$, with $f_X(x) = \mathbf{1}_{[0,1]}(x)$).

Problem 1. *Compute the joint density function of $X+Y$.*

Solution. For two independent random variables X and Y, the density function of their sum $Z = X+Y$ can be calculated using the convolution formula:

$$f_{X+Y}(z) = \int_{-\infty}^{+\infty} f_X(x)f_Y(z-x)\,dx$$
$$= \int_0^1 \mathbf{1}_{[0,1]}(x)\mathbf{1}_{[0,1]}(z-x)\,dx$$

Since $f_X(x) = \mathbf{1}_{[0,1]}(x)$ and $f_Y(y) = \mathbf{1}_{[0,1]}(y)$, the integration region must satisfy both $0\le x\le 1$ and $0 \le z-x \le 1$, which means $0\le x\le 1$ and $z-1\le x\le z$.

This can be divided into three cases:

$$f_{X+Y}(z) = \int_0^1 \mathbf{1}_{0\le z-x\le 1}\,dx = \begin{cases} \int_0^z 1\,dx = z, & \text{if } z\in[0,1] \\ \int_{z-1}^1 1\,dx = 2-z, & \text{if } z\in(1,2] \\ 0, & \text{otherwise} \end{cases}$$

Therefore, the density function of $X+Y$ is:

$$f_{X+Y}(z) = \begin{cases} z, & \text{if } z \in [0,1] \\ 2-z, & \text{if } z \in (1,2] \\ 0, & \text{otherwise} \end{cases}$$

This is a triangular distribution, reaching its maximum value of 1 at $z=1$.

Problem 2. *Let $X_1, \ldots, X_n$ be independent random variables, each uniformly distributed on $[0,1]$. Compute $F_{X_1+\ldots+X_n}(z)$ for $z \in [0,1]$.*

Solution. We will use induction to prove that $F_{X_1+\ldots+X_n}(z) = \frac{z^n}{n!}$ for $z \in [0,1]$.

Base case: When $n=1$, $F_{X_1}(z) = z$ for $z \in [0,1]$, which is the distribution function of a uniform distribution on $[0,1]$.

Induction hypothesis: Assume that for $n-1$, we have $F_{X_1+\ldots+X_{n-1}}(z) = \frac{z^{n-1}}{(n-1)!}$ for $z \in [0,1]$.

Induction step: We need to prove that $F_{X_1+\ldots+X_n}(z) = \frac{z^n}{n!}$ for $z \in [0,1]$.

Using the convolution formula and the induction hypothesis, we can calculate:

$$\begin{aligned} F_n(z) &= \int_{-\infty}^{+\infty} \int_{-\infty}^{z-x} f(x) f_{n-1}(y)\, dy\, dx \\ &= \int_{-\infty}^{+\infty} f(x) F_{n-1}(z-x)\, dx \end{aligned}$$

For $z \in [0,1]$, substituting the known conditions:

$$\begin{aligned} F_n(z) &= \int_0^z \frac{(z-x)^{n-1}}{(n-1)!}\, dx \\ &= \frac{1}{(n-1)!} \int_0^z (z-x)^{n-1}\, dx \end{aligned}$$

Using the substitution $u = z-x$, $dx = -du$, when $x=0$ we have $u=z$, and when $x=z$ we have $u=0$:

$$\begin{aligned} F_n(z) &= \frac{1}{(n-1)!} \int_z^0 u^{n-1}(-du) \\ &= \frac{1}{(n-1)!} \int_0^z u^{n-1}\, du \\ &= \frac{1}{(n-1)!} \cdot \frac{z^n}{n} \\ &= \frac{z^n}{n!} \end{aligned}$$

Therefore, we have proven that for $z \in [0,1]$, $F_{X_1+\ldots+X_n}(z) = \frac{z^n}{n!}$.

Note: This result is only valid for $z \in [0,1]$. For $z > 1$, the distribution function expression becomes more complex.

Problem 3. *Let $X_1, X_2, \ldots$ be independent random variables, each uniformly distributed on $[0,1]$. Define $N = \min\{n \in \mathbb{N} : X_1 + X_2 + \ldots + X_n > 1\}$. Compute $\mathbb{E}[N]$.*

Solution. We define $N = \min\{n \in \mathbb{N} : X_1 + X_2 + \ldots + X_n > 1\}$.

First, observe that the event $\{N \geq n\}$ is equivalent to the event $\{X_1 + X_2 + \ldots + X_{n-1} \leq 1\}$. This is because $N \geq n$ means that the sum of the first $n-1$ random variables is not yet sufficient to exceed 1.

Therefore:

$$\begin{aligned} P(N \geq n) &= P(X_1 + X_2 + \ldots + X_{n-1} \leq 1) \\ &= F_{X_1+\ldots+X_{n-1}}(1) \\ &= \frac{1^{n-1}}{(n-1)!} \\ &= \frac{1}{(n-1)!} \end{aligned}$$

Using the formula for the expectation of a discrete random variable $E[N] = \sum_{n=1}^{\infty} P(N \geq n)$, we have:

$$\begin{aligned} E[N] &= \sum_{n=1}^{\infty} P(N \geq n) \\ &= \sum_{n=1}^{\infty} \frac{1}{(n-1)!} \\ &= \sum_{n=0}^{\infty} \frac{1}{n!} \\ &= e \end{aligned}$$

The last step follows from the fact that $\sum_{n=0}^{\infty} \frac{1}{n!} = e$, which is the series expansion of the natural constant e.

Therefore, $E[N] = e \approx 2.71828...$

§1.21 Recitation (03-21)

Problem 1

Is is possible to have 2 biased dice, such that the sum is uniformly distributed in {2,3,...12}?

Problem 2

Flip a fair coin, What is the expected time to see the 1-st occur of HHT?

Problem 3

Let X be a non-negative r.v. show that

$$E[X^r] = \int_0^{+\infty} r x^{r-1} P(X > x) dx, r > 0$$

Problem 4

Gamma $(n, \lambda), f_{n,\lambda} = \frac{\lambda^n x^{n-1}}{P(n)} e^{-\lambda x}$,where $P(n) = (n-1)!$

Show that if $X_1, X_2, ..., X_n$ is independent $Exp(\lambda)$ then $X_1 + X_2 + \cdots + X_n \sim Gamma(n, \lambda)$

§1.22 Lecture 16 (03-24)

+5 Questions; Grade the best 4.
Bookwork Content ($\star$ - proofs are examinable)

1. Def. of algebra: algebra generated by class of subsets.
   Important examples from real line & Discrete sets. (identify the algebra/σ-algebra generated by given class of sets)*
2. $\cdots$ σ-algebra, σ-algebra $- - - - -$
   $- - - - - - - - - -$
3. Def. of Content. Subadditivity prop. Important examples from $\mathbb{R}$.
4. $\cdots$ Measure. Subadditivity*. Continuity from above/below.*
5. Lebesgue measure. Borel sets. examples of $\mathcal{B}(\mathbb{R})$. Lebesgue-Stieltjes measure.
6. Extension Thm.
7. Def. of π-system. Example of π-system that generates $\mathcal{B}(\mathbb{R})$. Uniqueness Thm. Application of Uniqueness Thm. to show the uniqueness of Lebesgue measure.*
8. Def. of mble funct./r.v., Borel funct. σ-algebra generated by r.v.
9. Equivalent cond. for a funct. being Borel.* Be able to prove certain funct. are m'ble.
   Operation of m'ble funct.*
10. Construction of Lebesgue integral (simple $\to$ non-negative Borel $\to$ general) $\{X < a, a \in \mathbb{R}\}$
11. Monotone Conv. Thm.
12. absolute Cont. and Radon-Nikodym Derivative.

Common distributions:
Bernoulli, Binomial, Poisson, Geometric, Uniform, Exponential, Gaussian

§1.23 Lecture 17 (04-07)

Random walk
Markov property

$$P[S_{n+m}|S_0, S_1, \cdots, S_m] = P[S_{n+m} = j|S_m]$$

"position after m-th step does not depend on the info before m"
Let

$$T_y^0 = 0$$
$$T_y^k = \inf\{n \geq T_y^{k-1} : S_n = y\}$$

Definition 1.23.1

y is recurrent if $P[T_y^k < \infty] = 1$ for all k
y is transient if $P[T_y^k < \infty] < 1$ for some k

Remark 1

If y is recurrent

$$P[T_y^k < \infty] = P[T_y^k < \infty | T_y^{k-1} < \infty] P[T_y^k < \infty] + P[T_y^k < \infty | T_y^{k-1} = \infty] P[T_y^k < \infty]$$