UNIVERSITY OF TORONTO | Statistical Sciences

DoSS Summer Bootcamp Probability Module 1

Ichiro Hashimoto

University of Toronto

July 8, 2024

Roadmap

A bridge connecting undergraduate probability and graduate probability

Undergraduate-level probability

- Concrete;
- Examples and scenarios;
- Rely on computation...

Roadmap

A bridge connecting undergraduate probability and graduate probability

Undergraduate-level probability

- Concrete;
- Examples and scenarios;
- Rely on computation...

Graduate-level probability

- Abstract (measure theory);
- Laws and properties;
- Rely on construction and inference...

Roadmap

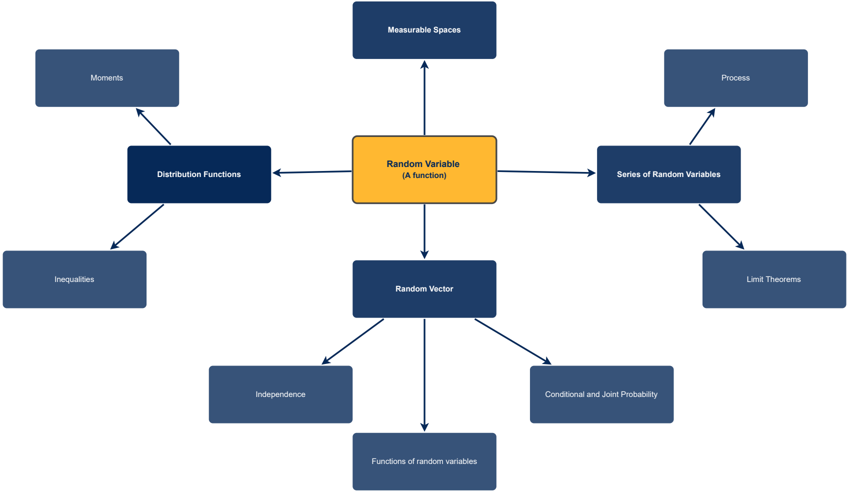


Figure: Roadmap

Outline

- Measurable spaces
  - ▷ Sample Space
  - ▷ σ-algebra
- Probability measures
  - ▷ Measures on σ-field
  - ▷ Basic results

Today

- Conditional probability
  - ▷ Bayes' rule
  - ▷ Law of total probability

→ Module 2.

Measurable spaces

Sample Space

The sample space Ω is the set of all possible outcomes of an experiment.

Examples:

- Toss a coin: $\{H, T\}$ $= \Omega$
- Roll a die: $\{1, 2, 3, 4, 5, 6\}$ $= \Omega$

Measurable spaces

Sample Space

The sample space Ω is the set of all possible outcomes of an experiment.

Examples:

- Toss a coin: $\{H, T\}$
- Roll a die: $\{1, 2, 3, 4, 5, 6\}$

Event

An event is a collection of possible outcomes (subset of the sample space).

Examples:

- Get head when tossing a coin: $\{H\}$ $\subset \{H, T\} = \Omega$
- Get an even number when rolling a die: $\{2, 4, 6\}$ $\subset \{1,2,3,4,5,6\} = \Omega$

ex1) Tossing a coin twice

$$\Omega = \{HH, HT, TH, TT\} \longrightarrow \text{discrete.}$$

$$\mathbb{P}(HH) = \mathbb{P}(HT) = \mathbb{P}(TH) = \mathbb{P}(TT) = \frac{1}{4}$$

Let $X =$ the number of H

$$\mathbb{P}(X=0) = \mathbb{P}(X=2) = 1/4$$

$$\mathbb{P}(X=1) = 1/2$$

$$\mathbb{P}(X=0) + \mathbb{P}(X=1) + \mathbb{P}(X=2) = 1$$

$$\mathbb{E}X = \frac{1}{4}\cdot 0 + \frac{1}{2}\cdot 1 + \frac{1}{4}\cdot 2 = 1$$

ex2) Let $X \sim N(\mu, \sigma^2)$ gaussian

continuous

Density $p(x) = \dfrac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\dfrac{(x-\mu)^2}{2\sigma^2}\right)$

$$\int_{-\infty}^{\infty} p(x)\,dx = 1$$

$$\mathbb{E}X = \int_{-\infty}^{\infty} x\,p(x)\,dx = \mu$$

Discrete case.

assume X only takes integer values for simplicity

$$\mathbb{P}(X \leq k) = \sum_{l \leq k} \mathbb{P}(X = l)$$

$$\mathbb{E}X = \sum_{k=-\infty}^{\infty} k\,\mathbb{P}(X = k)$$

Continuity case

$$\mathbb{P}(X \leq x) = \int_{-\infty}^{x} p(z)\,dz$$

$$\mathbb{E}X = \int_{-\infty}^{\infty} x\,p(x)\,dx$$

Question: Is there any way to explain them in a unified way?

Observation

If $A \cap B = \phi$, then $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$.

For a discrete case. $\{X = k\}$ are disjoint.

$$1 = \underbrace{\sum_{k=-\infty}^{\infty} \mathbb{P}(X = k)}_{\text{countable sum}}$$

But for continuous case,

$$\mathbb{P}(X = x) = 0 \quad \text{for any } x$$

Therefore.

$$1 = \underbrace{\sum_{x \in \mathbb{R}} \mathbb{P}(X = x)}_{\text{uncountable sum}} = \sum_{x \in \mathbb{R}} 0 = 0$$

contradiction?

$\Longrightarrow$ uncountable sum is problematic.

$\Longrightarrow$ let's focus on countable sum

Measurable spaces

Using (i) and (ii) $\Rightarrow$ If $A_1, \cdots \in \mathcal{F}$, $\bigcap_{i=1}^{\infty} A_i \in \mathcal{F}$

σ-algebra

A σ-algebra (σ-field) $\mathcal{F}$ on Ω is a non-empty collection of subsets of Ω such that

- (i) If $A \in \mathcal{F}$, then $A^c \in \mathcal{F}$, $\rightarrow$ complement is also in $\mathcal{F}$
- (ii) If $A_1, A_2, \cdots \in \mathcal{F}$, then $\cup_{i=1}^{\infty} A_i \in \mathcal{F}$. $\rightarrow$ Countable union of subsets of $\mathcal{F}$ is also in $\mathcal{F}$.

Remark: $\varnothing, \Omega \in \mathcal{F}$

(Pf) Let $A \in \mathcal{F}$.

By (i), $A^c \in \mathcal{F}$.

By (ii) $\underbrace{A \cup A^c}_{=\Omega} \in \mathcal{F}$ so $\Omega \in \mathcal{F}$.

By (i) again, $\Omega^c \in \mathcal{F}$. So, $\phi \in \mathcal{F}$ //.

UNIVERSITY OF TORONTO

Construction of Probability Theory

Outline.

1) Define the collection of subsets of Ω, $\mathcal{F}$ (σ-algebra) on which we can define "Probability measure".

2) Define probability measure as a function
$$\mathbb{P} : \mathcal{F} \longrightarrow [0, 1]$$
which has "countable additivity".

3) $(\Omega, \mathcal{F}, \mathbb{P})$ is called "Probability triple".

Ω: sample space

$\mathcal{F}$: σ-algebra

$\mathbb{P}$: probability measure

Probability measures

Measures on σ-field

A function $\mu : \mathcal{F} \to R^+ \cup \{+\infty\}$ is called a measure if

- $\mu(\varnothing) = 0$,
- If $A_1, A_2, \dots \in \mathcal{F}$ and $A_i \cap A_j = \varnothing$, then $\mu(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mu(A_i)$.

If $\mu(\Omega) = 1$, then μ is called a probability measure.

countable additivity

Probability measures

Measures on σ-field

A function $\mu : \mathcal{F} \to R^+ \cup \{+\infty\}$ is called a measure if

- $\mu(\varnothing) = 0$,
- If $A_1, A_2, \dots \in \mathcal{F}$ and $A_i \cap A_j = \varnothing$, then $\mu(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mu(A_i)$.

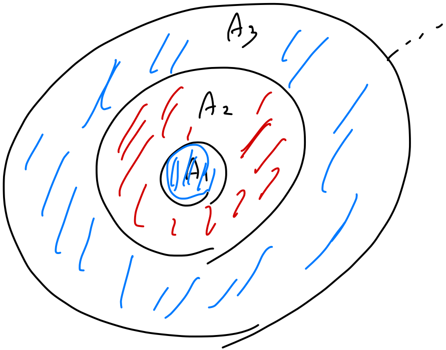
If $\mu(\Omega) = 1$, then μ is called a probability measure.

Properties:

- Monotonicity: $A \subseteq B \quad \Rightarrow \quad \mu(A) \leq \mu(B)$
- Subadditivity: $A \subseteq \cup_{i=1}^{\infty} A_i \quad \Rightarrow \quad \mu(A) \leq \sum_{i=1}^{\infty} \mu(A_i)$
- Continuity from below: $A_i \nearrow A \quad \Rightarrow \quad \mu(A_i) \nearrow \mu(A)$
- Continuity from above: $A_i \searrow A$ and $\mu(A_1) < \infty \quad \Rightarrow \quad \mu(A_i) \searrow \mu(A)$

Probability measures

Proof of continuity from below:



Let $A_i \in \mathcal{F}$, $A_1 \subset A_2 \subset \cdots$

$\bigcup_{i=1}^{\infty} A_i = A$.

Let $B_i = A_i \setminus A_{i-1}$, $i \geq 2$.

$B_1 = A_1$

Then B_i are disjoint.

$\bigcup_{i=1}^{\infty} B_i = \bigcup_{i=1}^{\infty} A_i = A$

By countable additivity $\mu(A) = \mu\left(\bigcup_{i=1}^{\infty} B_i\right) = \sum_{i=1}^{\infty} \mu(B_i)$

Note that $\underbrace{A_i = A_{i-1} \cup B_i}_{\text{disjoint union.}}$ implies

$$\mu(A_i) = \mu(A_{i-1}) + \mu(B_i)$$ by countable additivity.

Therefore $\mu(B_i) = \mu(A_i) - \mu(A_{i-1})$

Thus $$\sum_{i=1}^{n} \mu(B_i) = \mu(B_1) + \sum_{i=2}^{n} \left(\mu(A_i) - \mu(A_{i-1}) \right)$$

$$= \mu(A_1) + \mu(A_n) - \mu(A_1)$$

$$= \mu(A_n)$$

So, $$\sum_{i=1}^{\infty} \mu(B_i) = \lim_{n\to\infty} \mu(A_n)$$

Thus, $$\mu(A) = \lim_{n\to\infty} \mu(A_n)$$

Probability measures

$A_i \searrow A$

Proof of continuity from above:

$\mu(A_1) < \infty$, $A_1 \supset A_2 \supset A_3 \supset \cdots$, $\bigcap_{i=1}^{\infty} A_i = A$

$B_i = A_1 - A_i$

Then $B_1 \subset B_2 \subset B_3 \subset \cdots$, $\bigcup_{i=1}^{\infty} B_i = A_1 \setminus A$

By the continuity from below,

$$\lim_{n\to\infty} \mu(B_n) = \mu\left(\bigcup_{i=1}^{\infty} B_i\right) = \mu(A_1 \setminus A) = \mu(A_1) - \mu(A)$$

Remark: $\mu(A_i) < \infty$ is vital.

Note that $\mu(B_n) = \mu(A_1 \setminus A_n) \overset{\downarrow}{=} \mu(A_1) - \mu(A_n)$

Thus $\lim_{n\to\infty} \left(\mu(A_1) - \mu(A_n) \right) = \mu(A_1) - \mu(A)$ ← $\mu(A_1) < \infty$

$\therefore \quad \lim_{n\to\infty} \mu(A_n) = \mu(A)$

Summary $(\Omega, \mathcal{F}, \mathbb{P})$ probability triple.

"Countable additivity" is the key.

Question How can $(\Omega, \mathcal{F}, \mathbb{P})$ provides unified theory?

Observation

$X : \Omega \to \mathbb{R}$ random variable.

$$\Omega = \{ x \in \mathbb{R} \}$$

$$= \bigcup_{i=-\infty}^{\infty} \{ x \in [i, i+1) \}$$

By countable additivity

$$1 = \mathbb{P}(\Omega) = \sum_{i=-\infty}^{\infty} \mathbb{P}(X \in [i, i+1))$$

$$\Omega = \bigcup_{i=-\infty}^{\infty} \left\{ x \in \left[\frac{i}{n}, \frac{i+1}{n} \right) \right\}$$

becomes finer as $n \nearrow \infty$

$$1 = \mathbb{P}(\Omega) = \sum_{i=-\infty}^{\infty} \mathbb{P}\left(X \in \left[\frac{i}{n}, \frac{i+1}{n}\right)\right)$$

Approximation of Expectation

$$\mathbb{E}X \simeq \sum_{i=-\infty}^{\infty} \frac{i}{n} \cdot \mathbb{P}\left(X \in \left[\frac{i}{n}, \frac{i+1}{n}\right)\right)$$

should become more precise as $n \to \infty$

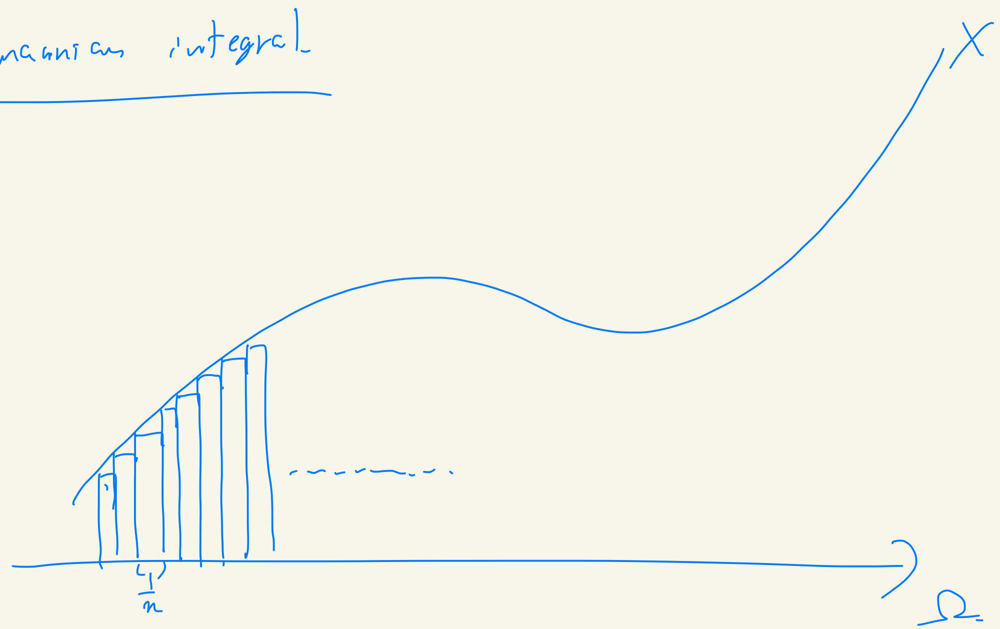
We can use this observation to define $\mathbb{E}X$

$$\mathbb{E}X = \lim_{n\to\infty} \sum_{i=-\infty}^{\infty} \frac{i}{n} \mathbb{P}\left(X \in \left[\frac{i}{n}, \frac{i+1}{n}\right)\right)$$

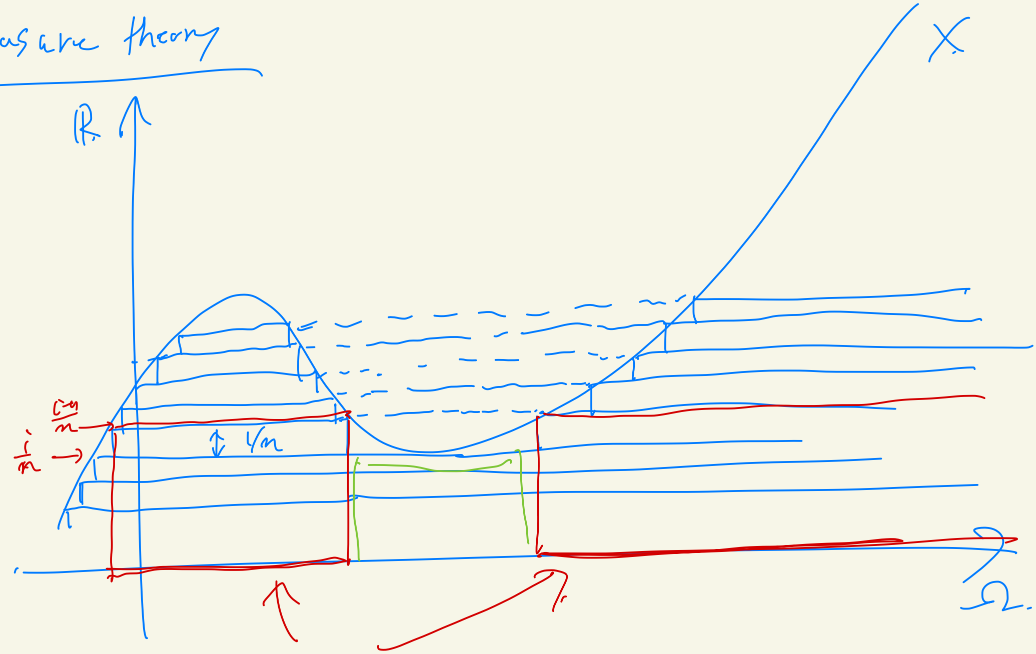
This looks similar to Riemannian integral

Difference between Riemannian integral

Riemannian integral



Measure theory



$\mathbb{P}\left(X \in \left[\frac{i}{n}, \frac{i+1}{n}\right)\right)$

$$\mathbb{E}X = \lim_{n\to\infty} \sum_{i=-\infty}^{\infty} \frac{i}{n} \mathbb{P}\left(X \in \left[\frac{i}{n}, \frac{i+1}{n}\right)\right) = \int_{\Omega} X \, d\mathbb{P}$$

We can show that

$$\mathbb{E}X = \sum_{i=1}^{\infty} k_i \mathbb{P}(X = k_i) \quad \text{discrete case.}$$

$$\mathbb{E}X = \int_{-\infty}^{\infty} x p(x) \, dx \quad \text{continuous case.}$$

Examples of σ-algebra (σ-field)

- Consider tossing a fair coin twice.

$$\Omega = \{HH, HT, TH, TT\}.$$

Trivial σ-algebra $\quad \mathcal{F} = \mathcal{P}(\Omega)$ power set

$=$ the collection of all subsets in Ω.

Note: For any Ω, $\mathcal{P}(\Omega)$ always gives a trivial σ-algebra.

Less trivial example:

union makes Ω.

$$\mathcal{F} = \{\phi, \{HH\}, \{HT, TH, TT\}, \Omega\}$$

complement

is generated σ-algebra by $\mathcal{A} = \{HH\}$

Q. Is less trivial σ-algebra always possible?

A. Easily construct such example by "generating" σ-algebra.

<u>Prop</u> Let $\mathcal{A}$ be a collection of subsets of Ω.

We always have the smallest σ-algebra $\mathcal{F}$ including $\mathcal{A}$; i.e. $\mathcal{A} \subset \mathcal{F}$.

Such $\mathcal{F}$ is denoted by $\mathcal{F} = \sigma(\mathcal{A})$ and called σ-algebra generated by $\mathcal{A}$.

(pf) Recall that $\mathcal{P}(\Omega)$ is a σ-algebra.

So it is valid to take the intersection of all σ-algebra containing $\mathcal{A}$.

Then define

$$\mathcal{F} = \bigcap \mathcal{G}.$$

$\mathcal{G}$: all σ-algebra containing $\mathcal{A}$.

Since intersection of σ-algebra is also σ-algebra, $\mathcal{F}$ defined above must be the smallest σ-algera containing $\mathcal{A}$. //

Q. How can we construct a meaningful σ-algebra on $\mathbb{R}$.

A. We can generated σ-algebra that contains all intervals.

(a, b)
$(a, b]$
$[a, b)$
$[a, b]$

$+$ complements, countable intersections, countable unions of above

$+$ // of above

⋮

This σ-algeba is called Borel sets and denoted by $\mathcal{R}$.

$(\mathbb{R}, \mathcal{R})$

Actually $\mathcal{R}$ contains

$$\left\{ \begin{array}{l} \left. \begin{array}{l} \text{all open sets} \\ \text{all closed sets} \end{array} \right\} + \begin{array}{l} \text{combination} \\ \text{of them} \end{array} \\ \\ \\ \end{array} \right.$$

$A \in \mathcal{R} \Leftrightarrow A$ is measurable.

Q. Is there any subset of $\mathbb{R}$ that is $\underbrace{\text{not a Borel set}}_{\text{(unmeasurable)}}$?

A. Yes or No depending what axiom you choose to develop Math theory.

If we allow axiom of choice, Yes we construct unmeasurable set.

Probability measures

Examples:

$\Omega = \{\omega_1, \omega_2, \cdots\}$, $A = \{\omega_{a_1}, \cdots, \omega_{a_i}, \cdots\} \Rightarrow \mu(A) = \sum_{j=1}^{\infty} \mu(\omega_{a_j})$.
Therefore, we only need to define $\mu(\omega_j) = p_j \geq 0$.
If further $\sum_{i=1}^{\infty} p_j = 1$, then μ is a probability measure.

- Toss a coin:

Define. $\mathbb{P}(H) = \mathbb{P}(T) = \frac{1}{2}$. Then $\mathbb{P}$ is a probability measure

- Roll a die:

Define $\mathbb{P}(1) = \mathbb{P}(2) = \mathbb{P}(3) = \cdots = \mathbb{P}(6) = \frac{1}{6}$

then $\mathbb{P}$ is a probability measure.

Conditional probability

Original problem:

- What is the probability of some event A?
- $P(A)$ is determined by our probability measure.

already know this information

New problem:

- Given that B happens, what is the probability of some event A?
- $P(A \mid B)$ is the conditional probability of the event A given B.

UNIVERSITY OF TORONTO

Conditional probability

Original problem:

- What is the probability of some event A?
- $P(A)$ is determined by our probability measure.

New problem:

- Given that B happens, what is the probability of some event A?
- $P(A \mid B)$ is the conditional probability of the event A given B.

Example:

- Roll a die: $P(\{2\} \mid \text{even number})$

Conditional probability

Definition

~~Bayes' rule~~

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}, \quad P(B) > 0$$

Remark: Does conditional probability $P(\cdot \mid B)$ satisfy the axioms of a probability measure?

→ Need to check definition of probability measure.

1) $\mathbb{P}(\phi \mid B) = \frac{\mathbb{P}(\phi \cap B)}{P(B)} = \frac{\mathbb{P}(\phi)}{P(B)} = \frac{0}{P(B)} = 0$

2) $\mathbb{P}(\Omega \mid B) = \frac{\mathbb{P}(\Omega \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B)}{\mathbb{P}(B)} = 1$

3) Let $A_1, A_2, \dots \in \mathcal{F}$, s.t. $A_i \cap A_j = \phi$.

Then
$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i \mid B \right) = \frac{\mathbb{P}\left(\left(\bigcup_{i=1}^{\infty} A_i \right) \cap B \right)}{\mathbb{P}(B)}$$

$$= \frac{\mathbb{P}\left(\bigcup_{i=1}^{\infty} (A_i \cap B) \right)}{\mathbb{P}(B)}$$

use countable additivity of $\mathbb{P}$.

$$= \frac{\sum_{i=1}^{\infty} \mathbb{P}(A_i \cap B)}{\mathbb{P}(B)}$$

$$= \sum_{i=1}^{\infty} \frac{\mathbb{P}(A_i \cap B)}{\mathbb{P}(B)}$$

used definition of conditional probability.

$$= \sum_{i=1}^{\infty} \mathbb{P}(A_i \mid B)$$

Conditional probability

Multiplication rule

$$P(A \cap B) = P(A \mid B)P(B) = P(B \mid A)P(A)$$

Generalization:

Law of total probability

Let $A_1, A_2, \cdots, A_n$ be a partition of ω, such that $P(A_i) > 0$, then

Ω

$$P(B) = \sum_{i=1}^{n} \underbrace{P(A_i)P(B \mid A_i)}_{= \, \mathbb{P}(A_i \cap B)}$$

Problem Set

Problem 1: Prove that for a σ-field $\mathcal{F}$, if $A_1, A_2, \cdots \in \mathcal{F}$, then $\cap_{i=1}^{\infty} A_i \in \mathcal{F}$.

Problem 2: Prove monotonicity and subadditivity of measure μ on σ-field.

Problem 3: (Monty Hall problem) Suppose you're on a game show, and you're given the choice of three doors: Behind one door is a car; behind the others, goats. You pick a door, say No. 1, and the host, who knows what's behind the doors, opens another door, say No. 3, which has a goat. He then says to you, "Do you want to pick door No. 2?" Is it to your advantage to switch your choice?
(Assumptions: the host will not open the door we picked and the host will only open the door which has a goat.)