UNIVERSITY OF TORONTO | Statistical Sciences

# DoSS Summer Bootcamp Probability Module 1

Ichiro Hashimoto

University of Toronto

July 11, 2023

# 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

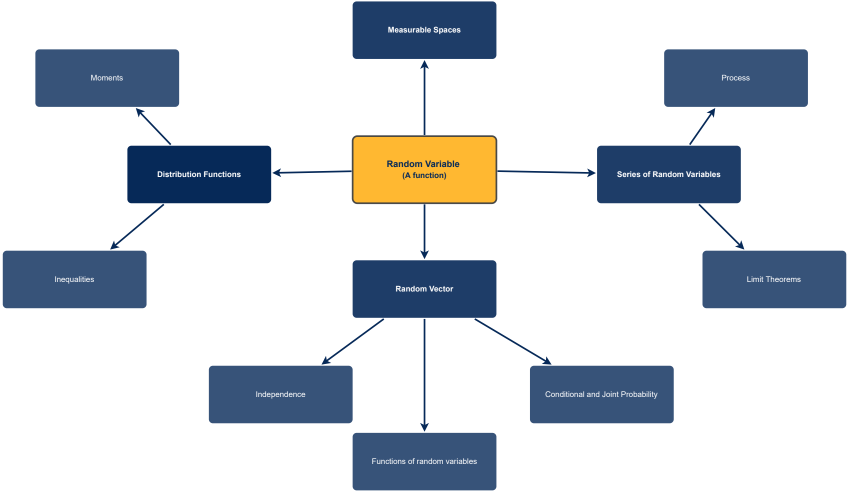



Figure: Roadmap

# Outline

- Measurable spaces
  - Sample Space
  - $\sigma$-algebra
- Probability measures
  - Measures on $\sigma$-field
  - Basic results
- Conditional probability
  - Bayes' rule
  - Law of total probability

# Measurable spaces

**Sample Space**

The sample space $\Omega$ is the set of all possible outcomes of an experiment.

**Examples:**

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

UNIVERSITY OF TORONTO

# Measurable spaces

**Sample Space**

The sample space $\Omega$ 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: $\{2, 4, 6\}$ $\subset \{1,2,3,4,5,6\} = \Omega$

If $|\Omega| = n$, then there $2^n$ events in total.

$\Omega = \{H, T\}$

$$\frac{\emptyset, \{H\}, \{T\}, \{H, T\}}{4 = 2^2}$$

$\Omega = \{1, 2, 3, 4, 5, 6\} \longrightarrow 2^6$ subsets

for each $i \in \Omega$. $\rightarrow$ $\dfrac{i \in A \text{ or } i \notin A}{2 \text{ choices for each } i}$

$\Longrightarrow 2^6$ subsets in total!

ex1) Tossing a coin twice

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

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

Let $X$ = the number of $H$

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

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

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

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

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

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

$$1 = \int_{-\infty}^{\infty} p(x)\,dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx$$

$$\mathbb{E}X = \int_{-\infty}^{\infty} x\,p(x)\,dx = \int_{-\infty}^{\infty} x\cdot\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx = \mu.$$

Discrete $$\mathbb{P}(X \le k) = \sum_{l=1}^{k} \mathbb{P}(X = l)$$

$$\mathbb{E}X = \sum_{k=1}^{\infty} k\,\mathbb{P}(X = k)$$

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

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

Question: Is there any way to explain those two in a unified manner?

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

(A, B are disjoint)

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

$$1 = \sum_{k=1}^{\infty} \mathbb{P}(X = k)$$ ← countable summation

But for continuous case,

$$\mathbb{P}(X = x) = 0 \quad x \in \mathbb{R}$$

Therefore

uncountable summation

$$1 \overset{?}{=} \sum_{x \in \mathbb{R}} \mathbb{P}(X = x) = \sum_{x \in \mathbb{R}} 0 \overset{?}{=} 0$$

$\Longrightarrow$ summation of uncountables doesn't work well

$\Longrightarrow$ might be better to focus countable sums.

# Construction of Probability theory

Outline.

1) Define the collection of subsets of $\Omega$, $\mathcal{F}$ ($\sigma$-algebra) on which we can "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"

sample space — $\Omega$

$\sigma$-algebra — $\mathcal{F}$

probability measure — $\mathbb{P}$

# Measurable spaces

## $\sigma$-algebra

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

- If $A \in \mathcal{F}$, then $A^c \in \mathcal{F}$, (i)
- If $A_1, A_2, \dots \in \mathcal{F}$, then $\underbrace{\cup_{i=1}^{\infty}}_{\text{Countable union}} A_i \in \mathcal{F}$. (ii)

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

(Proof)

Let $A \in \mathcal{F}$.

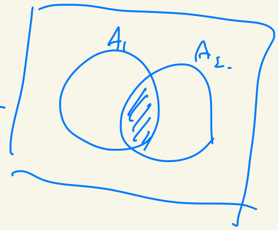
(i) $\Rightarrow$ $A^c \in \mathcal{F}$

(ii) $\Rightarrow$ $\underbrace{A \cup A^c}_{= \Omega} \in \mathcal{F}$ $\quad \therefore \Omega \in \mathcal{F}$

(i) $\Rightarrow$ $\phi = \Omega^c \in \mathcal{F}$.

- $\bigcap_{i=1}^{\infty} A_i \in \mathcal{F}$ $\qquad$ $\bigcap_{i=1}^{k} A_i = \bigcap_{i=1}^{\infty} A_i$

$A_i = \Omega$ for $i > k$




(Proof)

$$\bigcap_{i=1}^{\infty} A_i = \left( \bigcup_{i=1}^{\infty} A_i^c \right)^c$$

(i) $\Rightarrow A_i^c \in \mathcal{F}$.

(ii) $\Rightarrow \bigcup_{i=1}^{\infty} A_i^c \in \mathcal{F}$.

(i) $\Rightarrow \bigcap_{i=1}^{\infty} A_i = \left( \bigcup_{i=1}^{\infty} A_i^c \right)^c \in \mathcal{F}$. //

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

$\mathcal{F}$: $\sigma$-algebra generated by $\{HH\}$

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

---

$\mathbb{P}$ on $\mathcal{F}$ by $\quad \mathbb{P}(\phi) = 0, \ \mathbb{P}\{HH\} = \frac{1}{4}$

$\mathbb{P}\{HT, TH, TT\} = \frac{3}{4}, \ \mathbb{P}(\Omega) = 1$

$\{HH, HT\} \notin \mathcal{F}$

# Probability measures

$\phi = \{H\} \cap \{T\}$

## Measures on $\sigma$-field

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

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

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

countable additivity

# Probability measures

**Measures on $\sigma$-field**

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

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

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

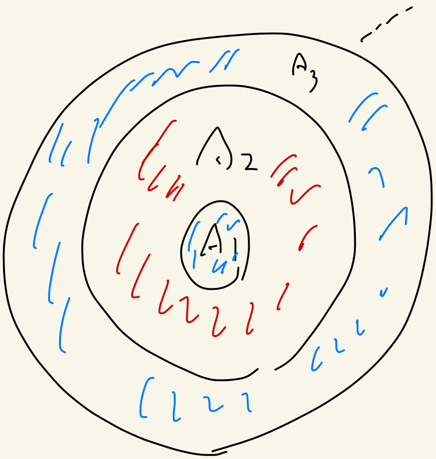
**Properties:**

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

Proof : Continuity from below.

If $A_i \in \mathcal{F}$, $A_1 \subset A_2 \subset A_3 \subset \cdots$

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



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

$B_1 = A_1$

Then $B_i$ are disjoint.

$$B_i = A_i \cap A_{i-1}^c \in \mathcal{F},$$

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

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

$$= \sum_{i=1}^{\infty} \mu(B_i) \quad \text{-----} \quad (*)$$

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

Therefore, $\sum_{i=1}^{k} \mu(B_i) = \sum_{i=2}^{k} \left(\mu(A_i) - \mu(A_{i-1})\right) + \mu(A_1) = \mu(A_k)$

That means, $(*)$ becomes

$$\mu(A) = \lim_{k\to\infty} \mu(A_k).$$

Continuity from above

$$\mu(A_1) < \infty, \quad A_1 \supset A_2 \supset A_3 \supset \cdots\cdots$$

$$A = \bigcap_{i=1}^{\infty} A_i$$

$$B_i = A_1 - A_i$$

Then $B_1 \subset B_2 \subset \cdots\cdots$

$$\bigcup_{i=1}^{\infty} B_i = A_1 \setminus A$$

By the continuity from below,

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

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

Note that $\mu(B_n) = \mu(A_1) - \mu(A_n)$

So $\lim_{n\to\infty} \{\mu(A_1) - \mu(A_n)\} = \mu(A_1) - \mu(A)$

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

# Probability measures

**Proof of continuity from below:**

# Probability measures

## Proof of continuity from above:

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

# 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 $\mu$ is a probability measure.

- Toss a coin:

- Roll a die:

# 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 $B$.

# 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

## 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?

# 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 $\omega$, such that $P(A_i) > 0$, then

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

# Problem Set

**Problem 1:** Prove that for a $\sigma$-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 $\mu$ on $\sigma$-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.)