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

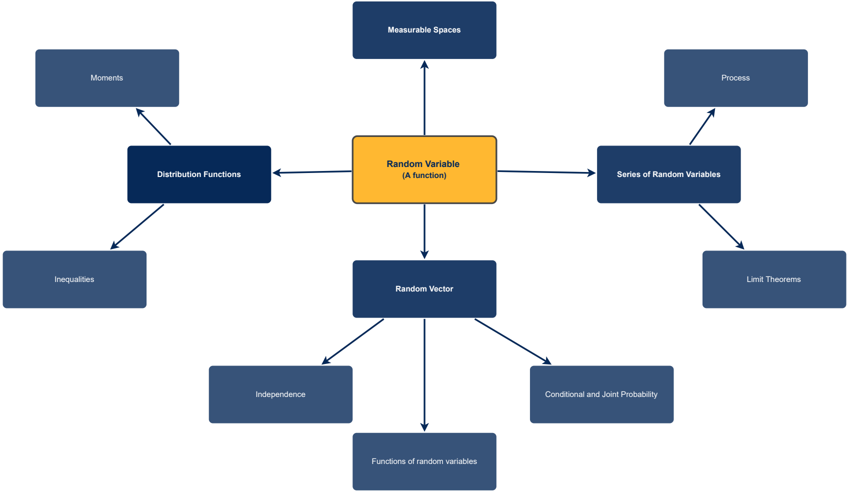


Figure: Roadmap

UNIVERSITY OF TORONTO

Outline

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

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:

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

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

UNIVERSITY OF TORONTO

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

$$\underbrace{\phi, \{H\}, \{T\}, \{H, T\}}_{4 = 2^2}$$

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

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

$\Longrightarrow 2^6$ subsets in total!

ex1) Tossing a coin twice

$\Omega = \{HH, HT, TH, TT\} \rightarrow$ 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 $\quad p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$

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

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

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

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

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

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

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

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

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

<u>Outline.</u>

1) Define the collection of subsets of Ω, $\mathcal{F}$ (σ-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"

Ω: sample space, $\mathcal{F}$: σ-algebra, $\mathbb{P}$: probability measure.

Measurable spaces

σ-algebra

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

- If $A \in \mathcal{F}$, then $A^c \in \mathcal{F}$, (i)
- If $A_1, A_2, \cdots \in \mathcal{F}$, then $\cup_{i=1}^{\infty} A_i \in \mathcal{F}$. (ii)

Countable union

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

(Proof)

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

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

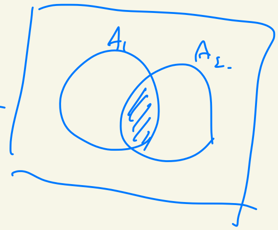
(ii) $\Rightarrow$ $\underbrace{A \cup A^c}_{= \Omega} \in \mathcal{F}$ $\therefore \Omega \in \mathcal{F}$

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

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

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

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

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

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

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

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

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

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

Probability measures

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

Measures on σ-field

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

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

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

disjoint

- $\mu(\varnothing) = 0$,
- 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.

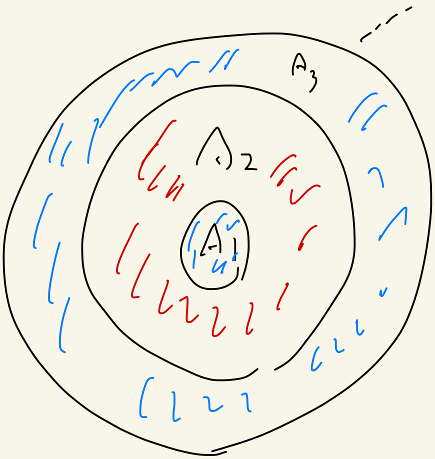
Properties:

- 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 $\mu(A_i) < \infty \quad \Rightarrow \quad \mu(A_i) \searrow \mu(A)$

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

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

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

Ω: sample space

$\mathcal{F}$: σ-algebra (σ-field)

$\mathbb{P}$: Probability measure.

"Countability" was the key.

1). Define σ-algebra $\mathcal{F}$ on which a probability can defined.

2) Define a probability measure $\mathbb{P}$ as $\mathbb{P} : \mathcal{F} \to [0, 1]$

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 implies

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

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

$$= \bigcup_{i=-\infty}^{\infty} \underbrace{\left\{ x \in \left[\frac{i}{n}, \frac{i+1}{n} \right) \right\}}_{\in \mathcal{F}?}$$

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

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

bridge between discrete probability and continuous.

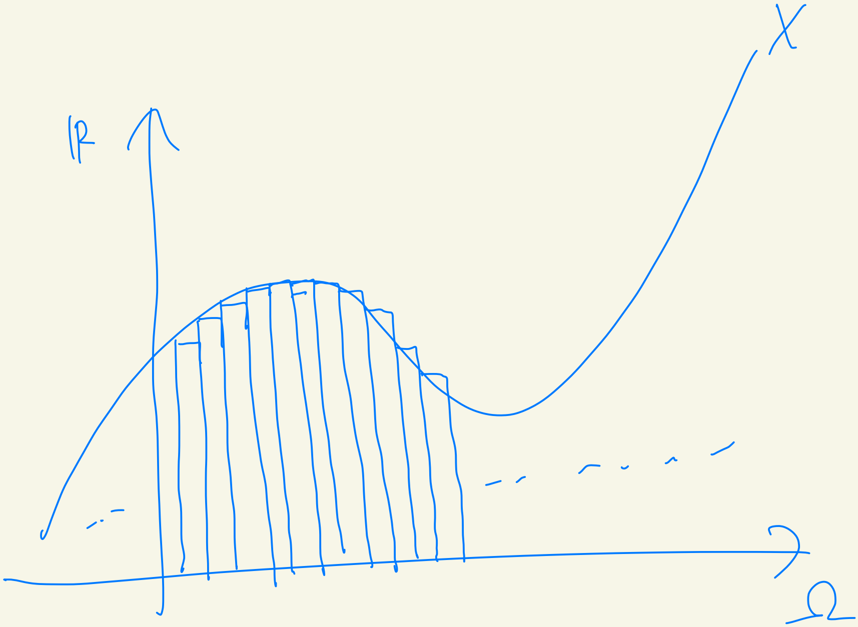
We can define EX from this observation

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

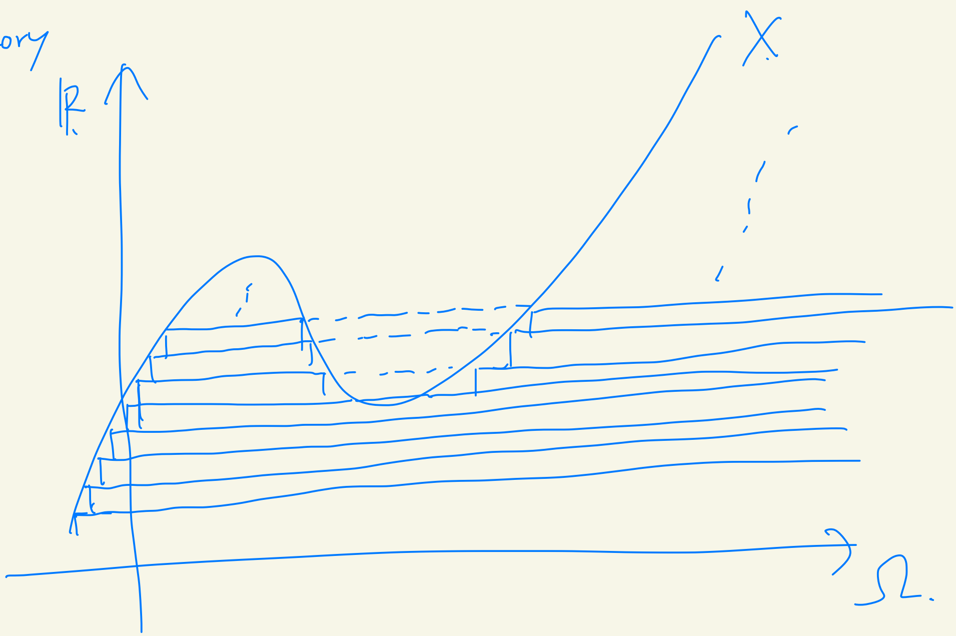
this is analogue to Riemannian sum for defining Riemannian integral.

Difference between Riemann and Lebesgue sum.

Riemann



Measure theory



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

We can show that

$$\begin{cases} EX = \displaystyle\sum_{i=1}^{\infty} h_i\, \mathbb{P}(X = h_i) & \text{for discrete case.} \\ EX = \displaystyle\int_{-\infty}^{\infty} x\, p(x)\, dx & \text{for continuous case.} \end{cases}$$

To make the above argument valid we need to choose appropriate $\mathcal{F}$.

Def (Borel sets).

Define a σ-algebra on $\mathbb{R}$ as "the smallest" σ-algebra that contains all intervals on $\mathbb{R}$.

We denote this σ-algebra by $\mathcal{R}$ or $\mathcal{B}$

$B \in \mathcal{R}$ is called a Borel set.

Then define $\mathcal{F}$ on Ω as

$$\mathcal{F} = \{ X^{-1}(B) : B \in \mathcal{R} \}$$

ensures $\{X \in [\frac{i}{n}, \frac{i+1}{n})\} \in \mathcal{F}$.

Remark $\mathcal{R}$. contains all intervals

+ (complements, countable union, countable intersection of intervals)

+ finite combination of these.

· $\mathcal{R}$ contains (any open sets, any closed sets, any single points, any countable sets.

$(\Omega, \mathcal{F}, \mathbb{P})$ $X: \Omega \to \mathbb{R}$.

We chose $\mathcal{F}$ so that $X^{-1}(B) \in \mathcal{F}$ for any $B \in \mathcal{R}$.

Def A function $X : \Omega \to \mathbb{R}$ is called a random variable if

$$\underbrace{X^{-1}(B) \in \mathcal{F} \quad \text{for any } B \in \mathcal{R}}.$$

We also say that X is measurable.

$$\mathbb{P}(X \in A) = \mathbb{P}\left(X^{-1}(A)\right) \qquad A \in \mathcal{R}.$$

Let us view this as a function from $\mathcal{R} \to [0,1]$

i.e. $\mu(A) = \mathbb{P}\left(X^{-1}(A)\right)$

Then μ is a probability measure on $\mathcal{R}$.

In other words, through X, a new probability measure is induced on $\mathcal{R}$.

We call μ a probability measure induced by X, $(\Omega, \mathcal{F}, \mathbb{P})$

$$(\Omega, \mathcal{F}, \mathbb{P}) \xrightarrow{X} \underbrace{(\mathbb{R}, \mathcal{R}, \mu)}_{\text{new triple is induced.}}$$

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 μ 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 $\cancel{\varpi}$, such that $P(A_i) > 0$, then

Ω

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

Problem Set

Problem 1: Prove that for a σ-field $\mathcal{F}$, if $A_1, A_2, \dots \in \mathcal{F}$, 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.)

UNIVERSITY OF TORONTO