



UNIVERSITY OF
TORONTO

Statistical Sciences

DoSS Summer Bootcamp Probability Module 2

Ichiro Hashimoto

University of Toronto

July 9, 2025

Recap

Learnt in last module:

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

Outline

- Independence of events
 - ▷ Pairwise independence, mutual independence
 - ▷ Conditional independence
- Random variables
- Distribution functions
- Density functions and mass functions
- Independence of random variables

Independence of events

Recall the Bayes rule:

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

- What if B does not change our belief about A ?

Independence of events

Recall the Bayes rule:

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

- What if B does not change our belief about A ?
- This means $P(A \mid B) = P(A)$.

Independence of events

Recall the Bayes rule:

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

- What if B does not change our belief about A ?
- This means $P(A | B) = P(A)$.
- Equivalently, $P(A \cap B) = P(A)P(B)$.

$$P(A|B) = P(A) \Leftrightarrow \frac{P(A \cap B)}{P(B)} = P(A)$$

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

Independence of events

Recall the Bayes rule:

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

- What if B does not change our belief about A ?
- This means $P(A | B) = P(A)$.
- Equivalently, $P(A \cap B) = P(A)P(B)$.

Independence of two events

Two events A and B are independent if $P(A \cap B) = P(A)P(B)$.

Remark:

Independence of events

Consider more than 2 events:

Pairwise independence

We say that events A_1, A_2, \dots, A_n are pairwise independent if

$$P(A_i \cap A_j) = P(A_i) \cdot P(A_j), \quad \forall i \neq j$$

Independence of events

Consider more than 2 events:

Pairwise independence

We say that events A_1, A_2, \dots, A_n are pairwise independent if

$$P(A_i \cap A_j) = P(A_i) \cdot P(A_j), \quad \forall i \neq j$$

(Mutual) independence

We say that events A_1, A_2, \dots, A_n are mutually independent or independent if for all subsets $I \in \{1, 2, \dots, n\}$

$$\underline{P(\cap_{i \in I} A_i)} = \prod_{i \in I} P(A_i)$$

Independence of events

Consider more than 2 events:

Pairwise independence

We say that events A_1, A_2, \dots, A_n are pairwise independent if

$$P(A_i \cap A_j) = P(A_i) \cdot P(A_j), \quad \forall i \neq j$$

Mutual independence

We say that events A_1, A_2, \dots, A_n are mutually independent or independent if for all subsets $I \in \{1, 2, \dots, n\}$

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

Remark:

mutual independence \Rightarrow pairwise independence is true
but \Leftarrow does not hold.

Independence of events

Example:

- Toss two fair coins.;
- $A = \{ \text{First toss is head} \}$, $B = \{ \text{Second toss is head} \}$, $C = \{ \text{Outcomes are the same} \}$;
- $A = \{HH, HT\}$, $B = \{HH, TH\}$, $C = \{HH, TT\}$;

Independence of events

Example:

- Toss two fair coins.;
- $A = \{ \text{First toss is head} \}$, $B = \{ \text{Second toss is head} \}$, $C = \{ \text{Outcomes are the same} \}$;
- $A = \{HH, HT\}$, $B = \{HH, TH\}$, $C = \{HH, TT\}$;
- $P(A \cap B) = P(A)P(B)$, $P(A \cap C) = P(A)P(C)$, $P(B \cap C) = P(B)P(C)$;

$$A \cap B = \{HH\}, \quad P(A \cap B) = \frac{1}{4}$$

$$A \cap C = \{HH\}, \quad P(A \cap C) = \frac{1}{4}$$

$$B \cap C = \{HH\}, \quad P(B \cap C) = \frac{1}{4}$$

A, B, C are pairwise independent.

Independence of events

Example:

- Toss two fair coins.;
- $A = \{ \text{First toss is head} \}$, $B = \{ \text{Second toss is head} \}$, $C = \{ \text{Outcomes are the same} \}$;
- $A = \{HH, HT\}$, $B = \{HH, TH\}$, $C = \{HH, TT\}$;
- $P(A \cap B) = P(A)P(B)$, $P(A \cap C) = P(A)P(C)$, $P(B \cap C) = P(B)P(C)$;
- $P(A \cap B \cap C) \neq P(A)P(B)P(C)$.

$$\overset{\sim \frac{1}{4}}{A \cap B \cap C} = \{HH\} \quad = \frac{1}{8}$$

Independence of events

Conditional independence

Two events A and B are conditionally independent given an event C if

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

definition is replaced by conditional probability
 $P(\cdot \mid C)$

Independence of events

Conditional independence

Two events A and B are conditionally independent given an event C if

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

Example:

Previous example continued:

- $A = \{HH, HT\}, B = \{HH, TH\}, C = \{HH, TT\};$
- $P(A \cap B \mid C) = ?, P(A \mid C)P(B \mid C) = ?$

$$P(A \cap B \mid C) = \frac{P(A \cap B \cap C)}{P(C)} = \frac{\frac{1}{4}}{\frac{1}{2}} = \frac{1}{2}$$

$$P(A \mid C) = \frac{P(A \cap C)}{P(C)} = \frac{\frac{1}{4}}{\frac{1}{2}} = \frac{1}{2}, \quad P(B \mid C) = \frac{1}{2}.$$

$$P(A \cap B \mid C) \neq P(A \mid C)P(B \mid C)$$

A and B are not conditionally independent

Independence of events

Conditional independence

Two events A and B are conditionally independent given an event C if

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

Example:

Previous example continued:

- $A = \{HH, HT\}, B = \{HH, TH\}, C = \{HH, TT\};$
- $P(A \cap B \mid C) = ?, P(A \mid C)P(B \mid C) = ?$

Remark:

Equivalent definition:

$$\underline{P(A \mid B, C) = P(A \mid C)}.$$

Random variables

Idea:

Instead of focusing on each events themselves, sometimes we care more about functions of the outcomes.

Random variables

Idea:

Instead of focusing on each events themselves, sometimes we care more about functions of the outcomes.

Example:

- Toss a fair coin twice: $\{HH, HT, TH, TT\}$
- Care about the number of heads: $\{2, 1, 0\}$

Random variables

Idea:

Instead of focusing on each events themselves, sometimes we care more about functions of the outcomes.

Example:

- Toss a fair coin twice: $\{HH, HT, TH, TT\}$
- Care about the number of heads: $\{2, 1, 0\}$

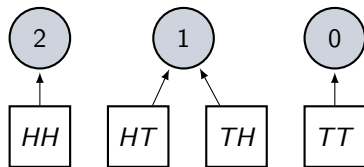


Figure: Mapping from the sample space to the numbers of heads

Random Variables

Example:

- Select twice from red and black ball with replacement: $\{RR, RB, BR, BB\}$
- Care about the number of red balls: $\{2, 1, 0\}$

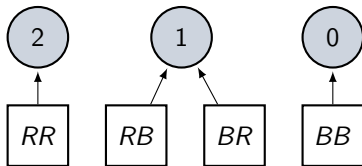


Figure: Mapping from the sample space to the numbers of red balls

Random Variables

Merits:

- Mapping the complicated events on σ -field to some numbers on real line.
- Simplify different events into the same structure

Random Variables

Merits:

- Mapping the complicated events on σ -field to some numbers on real line.
- Simplify different events into the same structure

Random Variables

Consider sample space Ω and the corresponding σ -field \mathcal{F} , for $X : \Omega \rightarrow \mathbb{R}$, if

$$\underline{A \in \mathcal{R} \quad (\text{Borel sets on } \mathbb{R}) \Rightarrow X^{-1}(A) \in \mathcal{F},}$$

then we call X as a random variable.

Here $X^{-1}(A) = \{\omega : X(\omega) \in A\}$.

We can also say X is \mathcal{F} -measurable.

Distribution functions

$$(\Omega, \mathcal{F}, \mathbb{P}) \xrightarrow{X} (\mathbb{R}, \mathcal{R}, \mu_X)$$

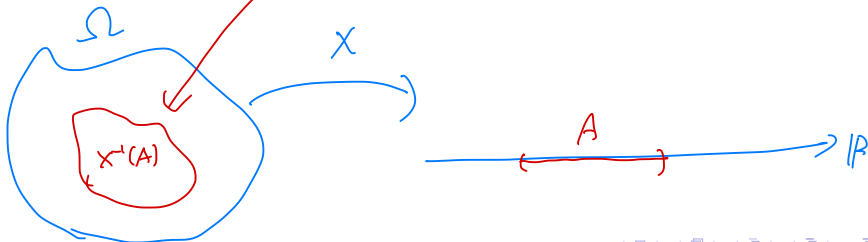
Probability measure $P(\cdot)$ on \mathcal{F} can induce a measure $\mu(\cdot)$ on \mathcal{R} :

Probability measure on \mathcal{R}

We can define a probability μ on $(\mathbb{R}, \mathcal{R})$ as follows:

$$\forall A \in \mathcal{R}, \quad \mu(A) := P(X^{-1}(A)) = P(X \in A).$$

Then μ is a probability measure and it is called the distribution of X .



Distribution functions

Probability measure $P(\cdot)$ on \mathcal{F} can induce a measure $\mu(\cdot)$ on \mathcal{R} :

Probability measure on \mathcal{R}

We can define a probability μ on (R, \mathcal{R}) as follows:

$$\forall A \in \mathcal{R}, \quad \mu(A) := P(X^{-1}(A)) = P(\underline{X \in A}).$$

Then μ is a probability measure and it is called the distribution of X .

Remark:

Verify that μ is a probability measure.

- $\mu(\mathbb{R}) = 1$.
- If $A_1, A_2, \dots \in \mathcal{R}$ are disjoint, then $\mu(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mu(A_i)$.

$$\begin{aligned} & \{X \in A\} \\ &= \{\omega \in \Omega : X(\omega) \in A\} \end{aligned}$$

Distribution functions

Consider the special set that belongs to \mathcal{R} , $(-\infty, x]$:

Cumulative Distribution Function

The cumulative distribution function of random variable X is defined as follows:

$$\underline{F(x) := P(X \leq x) = P(X^{-1}((-\infty, x]))}, \quad \forall x \in \mathbb{R}.$$

$$\begin{array}{c} \parallel \\ \mu(-\infty, x] \end{array}$$

Distribution functions

Consider the special set that belongs to \mathcal{R} , $(-\infty, x]$:

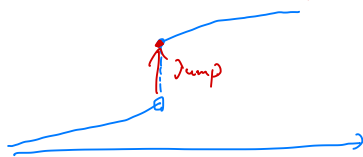
Cumulative Distribution Function

The cumulative distribution function of random variable X is defined as follows:

$$F(x) := P(X \leq x) = P(X^{-1}((-\infty, x])), \quad \forall x \in \mathbb{R}.$$

Properties of CDF:

- $\lim_{x \rightarrow \infty} F(x) = 1, \lim_{x \rightarrow -\infty} F(x) = 0$ *← can show by continuity of probability measure.*
- $F(\cdot)$ is non-decreasing
- $F(\cdot)$ is right-continuous *←*
- Let $F(x^-) = \lim_{y \nearrow x} F(y)$, then $F(x^-) = P(X < x)$
- $P(X = x) = F(x) - F(x^-)$ *← > 0 if there is a jump at x*



Distribution functions

Proofs of properties of CDF (first 2 properties):

$$F(x) = P(X \leq x)$$

When $x \nearrow \infty$ $\{X \leq x\} \nearrow \Omega$.

By the continuity from below,

$$\lim_{x \rightarrow \infty} P(X \leq x) = P(\Omega) = 1.$$

When $x \searrow -\infty$, $\{X \leq x\} \searrow \emptyset$.

By the continuity from above,

$$\lim_{x \rightarrow -\infty} P(X \leq x) = P(\emptyset) = 0.$$

Let $x_1 \leq x_2$.

Then $\{x \leq x_1\} \subset \{x \leq x_2\}$

$$\text{So, } \{x \leq x_2\} = \{x \leq x_1\} \cup \{x_1 < x \leq x_2\}$$


disjoint.

By countable additivity

$$F(x_2) = P(x \leq x_2) = P(x \leq x_1) + \underbrace{P(x_1 < x \leq x_2)}_{\geq 0}$$

$$\geq P(x \leq x_1) = F(x_1).$$

Density functions and mass functions

Classification of the random variables:

- Discrete random variable: X takes either a finite or countable number of possible numbers.
- Continuous random variable: The CDF is continuous everywhere.

Density functions and mass functions

Classification of the random variables:

- Discrete random variable: X takes either a finite or countable number of possible numbers.
- Continuous random variable: The CDF is continuous everywhere.

Another perspective (function):

- Discrete random variable: focus on the probability assigned on each possible values
- Continuous random variable: consider the derivative of the CDF (The continuous monotone CDF is differentiable almost everywhere)

Density functions and mass functions *Discrete case*

Probability mass function

The probability mass function of X at some possible value x is defined by

$$p_X(x) = P(X = x).$$

Relationship between PMF and CDF:

$$F(x) = P(X \leq x) = \sum_{y \leq x} p_X(y)$$



this sum is valid

since this is countable sum

Density functions and mass functions

Probability mass function

The probability mass function of X at some possible value x is defined by

$$p_X(x) = P(X = x).$$

Relationship between PMF and CDF:

$$F(x) = P(X \leq x) = \sum_{y \leq x} p_X(y)$$

Example:

Toss a coin

Density functions and mass functions

continuous

Probability density function

The probability density function of X at some possible value x is defined by

$$f_X(x) = \frac{d}{dx}F(x).$$

Relationship between PDF and CDF:

$$F(x) = P(X \leq x) = \int_{y \leq x} f_X(y) dy = \int_{-\infty}^x f_X(y) dy$$

Density functions and mass functions

Probability density function

The probability density function of X at some possible value x is defined by

$$f_X(x) = \frac{d}{dx}F(x).$$

Relationship between PDF and CDF:

$$F(x) = P(X \leq x) = \int_{y \leq x} f_X(y) dy = \int_{-\infty}^x f_X(y) dy$$

Example:

Independence of random variables

Define independence of random variables based on independence of events:

Independence of random variables

Suppose X_1, X_2, \dots, X_n are random variables on (Ω, \mathcal{F}, P) , then

def X_1, X_2, \dots, X_n are independent

$\Leftrightarrow \{X_1 \in A_1\}, \{X_2 \in A_2\}, \dots, \{X_n \in A_n\}$ are independent, $\forall A_i \in \mathcal{R}$

$$\Leftrightarrow P(\cap_{i=1}^n \{X_i \in A_i\}) = \prod_{i=1}^n P(\{X_i \in A_i\})$$

Independence of random variables

Example:

Toss a fair coin twice, denote the number of heads of the i -th toss as X_i , then X_1 and X_2 are independent.

- A_i can be $\{0\}$ or $\{1\}$
- $\{(0, 0), (0, 1), (1, 0), (1, 1)\}$
- $P(\{X_1 \in A_1\} \cap \{X_2 \in A_2\}) = \frac{1}{4}$
- $P(\{X_1 \in A_1\}) = P(\{X_2 \in A_2\}) = \frac{1}{2}$

$$\left. \begin{array}{l} \\ \\ \\ \end{array} \right) \quad P(\{X_1 \in A_1\} \cap \{X_2 \in A_2\}) \\ = P(X_1 \in A_1) \times P(X_2 \in A_2)$$

Independence of random variables

Example:

Toss a fair coin twice, denote the number of heads of the i -th toss as X_i , then X_1 and X_2 are independent.

- A_i can be $\{0\}$ or $\{1\}$
- $\{(0, 0), (0, 1), (1, 0), (1, 1)\}$
- $P(\{X_1 \in A_1\} \cap \{X_2 \in A_2\}) = \frac{1}{4}$
- $P(\{X_1 \in A_1\}) = P(\{X_2 \in A_2\}) = \frac{1}{2}$

Remark:

How to check independence in practice?

Independence of random variables

Corollary of independence

If X_1, \dots, X_n are random variables, then X_1, X_2, \dots, X_n are independent if

$$P(X_1 \leq x_1, \dots, X_n \leq x_n) = \prod_{i=1}^n P(X_i \leq x_i)$$

A_i 's are replaced by $[-\infty, x_i]$

Independence of random variables

Corollary of independence

If X_1, \dots, X_n are random variables, then X_1, X_2, \dots, X_n are independent if

$$P(X_1 \leq x_1, \dots, X_n \leq x_n) = \prod_{i=1}^n P(X_i \leq x_i)$$

Remark:

Independence of discrete random variables

Suppose X_1, \dots, X_n can only take values from $\{a_1, \dots\}$, then X_i 's are independent if

$$P(\cap \{X_i = a_i\}) = \prod_{i=1}^n P(X_i = a_i)$$

Problem Set

Problem 1: Give an example where the events are pairwise independent but not mutually independent.

Problem 2: Verify that the measure $\mu(\cdot)$ induced by $P(\cdot)$ is a probability measure on \mathcal{R} .

Problem 3: Prove properties 3 - 5 of CDF $F(\cdot)$.

Problem 4: Bob and Alice are playing a game. They alternatively keep tossing a fair coin and the first one to get a H wins. Does the person who plays first have a better chance at winning?