



UNIVERSITY OF
TORONTO

Statistical Sciences

DoSS Summer Bootcamp Probability Module 2

Ichiro Hashimoto

University of Toronto

July 9, 2026

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:

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

- 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, $P(A \cap B) = P(A)P(B)$.

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 ? *interpretation*
- 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 $P(A \cap B) = P(A)P(B)$. *definition*

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 $I \in \{1, 2, \dots, n\}$

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

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:

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

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

This confirms A and B are independent

Independence of events

Example:

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

$$P(A \cap B \cap C) = P(HH) = \left(\frac{1}{4}\right)$$

However, $P(A)P(B)P(C) = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \left(\frac{1}{8}\right)$

A, B, C are not independent

Independence of events

Conditional independence

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

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

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 given C .

Independence of events

Conditional independence

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

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

Example:

Previous example continued:

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

Remark:

Equivalent definition:

$$P(A | B, C) = P(A | C).$$

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: $\{2, 1, 0\}$

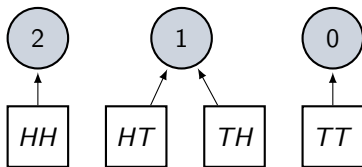


Figure: Mapping from the sample space to the numbers of heads

Random Variables

Example:

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

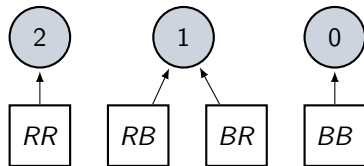


Figure: Mapping from the sample space to the numbers of red balls

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

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

then we call X as a random variable. *Q. Why do we need this?*

Here $X^{-1}(A) = \{\omega : X(\omega) \in A\}$. *A. We often want to consider.*

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

$$P(X \in \underline{I})$$

interval

To make this valid, we must have $\{X \in I\} \in \mathcal{F}$

since P is only defined on $\mathcal{F}$.

Distribution functions

Probability measure $P(\cdot)$ on $\mathcal{F}$ can induce a measure $\mu(\cdot)$ on $\mathcal{R}$ through r.v. X (Borel σ -algebra)

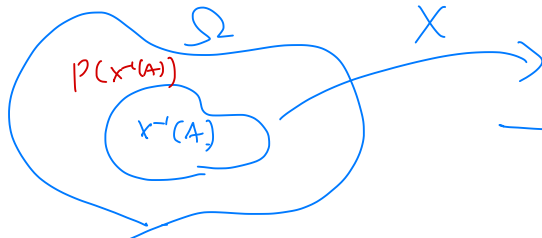
Probability measure on $\mathcal{R}$

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

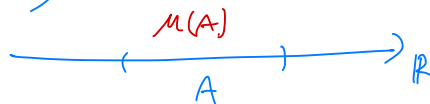
$$\forall A \in \mathcal{R}, \quad \mu(A) := P(X^{-1}(A)) = P(X \in A).$$

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

$$P: \mathcal{F} \rightarrow [0, 1]$$



$$\mu: \mathcal{R} \rightarrow [0, 1]$$



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:

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

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

Remark:

Verify that μ is a probability measure.

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

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = P\left(X \in \bigcup_{i=1}^{\infty} A_i\right) = P\left(\bigcup_{i=1}^{\infty} \{X \in A_i\}\right) = \sum_{i=1}^{\infty} P(X \in A_i) = \sum_{i=1}^{\infty} \mu(A_i)$$

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:

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

$$= \mu((-\infty, x])$$

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:

- $\lim_{x \rightarrow \infty} F(x) = 1$, $\lim_{x \rightarrow -\infty} F(x) = 0$
- $F(\cdot)$ is non-decreasing
- $F(\cdot)$ is right-continuous
- Let $F(x^-) = \lim_{y \nearrow x} F(y)$, then $F(x^-) = P(X < x)$
- $P(X = x) = F(x) - F(x^-)$

Distribution functions

Proofs of properties of CDF (first 2 properties):

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

$$\{x \leq x\} \nearrow \{x \in \mathbb{R}\} = \Omega.$$

→ continuity from below

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

$$\{x \leq x\} \searrow \emptyset$$

→ continuity from above.

Let $x_1 \leq x_2$.

$$\text{Then } F(x_2) - F(x_1) = P(X \leq x_2) - P(X \leq x_1)$$

$$= P(\{X \leq x_1\} \cup \{x_1 < X \leq x_2\}) - P(X \leq x_1)$$

disjoint

$$= \cancel{P(X \leq x_1)} + P(x_1 < X \leq x_2) - \cancel{P(X \leq x_1)} = P(x_1 < X \leq x_2) \geq 0$$

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

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: *for discrete r.v.*

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

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

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

Independence of events $A_1, \dots, A_n$

$$P\left(\bigcap_{i \in I} A_i\right) = \prod_{i \in I} P(A_i)$$

for any $I \subset \{1, \dots, n\}$

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 $(\Omega, \mathcal{F}, P)$, then

$X_1, X_2, \dots, X_n$ are independent

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

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

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

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

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

Instead of considering all $A_i \in \mathcal{R}$,
we only need to consider $A_i = (-\infty, x_i]$.

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?