UNIVERSITY OF TORONTO | Statistical Sciences

DoSS Summer Bootcamp Probability Module 2

Ichiro Hashimoto

University of Toronto

July 10, 2024

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?

= what if B is independent of A ?

UNIVERSITY OF TORONTO

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 $\underbrace{P(A \mid B) = P(A)}_{\text{knowing } B \text{ does not change the probability}}$.

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

use this as the definition of independence

UNIVERSITY OF TORONTO

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 $P(A \mid B) = P(A)$.
- 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 $A_1, A_2, \cdots, A_n$ 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 $A_1, A_2, \cdots, A_n$ 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 $A_1, A_2, \cdots, A_n$ are mutually independent or independent if for all subsets $I \in \{1, 2, \cdots, n\}$

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

Mutual independence $\Longrightarrow$ pairwise independence.

Choose $I = \{i, j\}$, $i \neq j$.

Then $\mathbb{P}(A_i \cap A_j) = \mathbb{P}(A_i)\,\mathbb{P}(A_j)$ by definition of independence.

Independence of events

Consider more than 2 events:

Pairwise independence

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

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

Remark:

Independence of events

Example:

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

Independence of events

Example:

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

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

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

$A \cap B = \{HH\}$

$B \cap C = \{HH\}$

$C \cap A = \{HH\}$

$\mathbb{P}(A \cap B) = \mathbb{P}(B \cap C) = \mathbb{P}(C \cap A) = \frac{1}{4} = \left(\frac{1}{2}\right)^2.$

A, B, C are pairwise independent.

Independence of events

Example:

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

$A \cap B \cap C = \{HH\} \Rightarrow \mathbb{P}(A \cap B \cap C) = \frac{1}{4} \neq \left(\frac{1}{2}\right)^3 = \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C)$

A, B, C are not mutually independent.

UNIVERSITY OF TORONTO

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:

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

$A \cap B = \{HH\}$,

$A \cap B \cap C = \{HH\}$,

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

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

So

$\mathbb{P}(A \cap B \mid C) = \frac{1}{2} \neq \left(\frac{1}{2}\right)^2$

$= \mathbb{P}(A \mid C) \times \mathbb{P}(B \mid C)$

Thus, 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

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

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:

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

can show this easily by algebra.

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

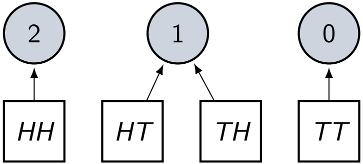


Figure: Mapping from the sample space to the numbers of heads

UNIVERSITY OF TORONTO

Random Variables

Example:

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

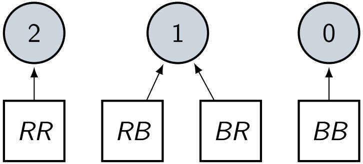


Figure: Mapping from the sample space to the numbers of red balls

Random Variables

Merits:

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

Borel sets

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 $X : \Omega \to \mathbb{R}$, if

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

then we call X as a random variable.
Here $X^{-1}(A) = \{\overset{\in \Omega}{\omega} : X(\omega) \in A\}$.
We can also say X is $\mathcal{F}$-measurable.

Distribution functions $(\Omega, \mathcal{F}, \mathbb{P}) \xrightarrow{X} (\mathbb{R}, \mathcal{R}, ?)$

probability measure?

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:

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

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

you can check

(i) $\mu(\phi) = 0$

(ii) If $A_1, A_2, \cdots$ are disjoint, $\mu(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mu(A_i)$

(iii) $\mu(\mathbb{R}) = 1$

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.

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

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:

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

$$\{X \leq x\} = \{X \in (-\infty, x]\}$$

$$= X^{-1}((-\infty, x])$$

Distribution functions

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

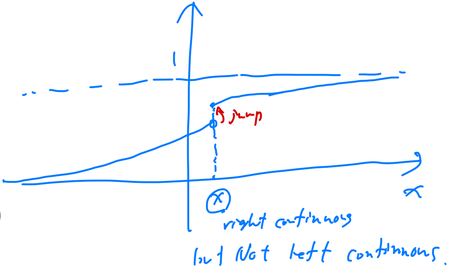
Cumulative Distribution Function

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:

- $\lim_{x\to\infty} F(x) = 1$, $\lim_{x\to-\infty} F(x) = 0$
- $F(\cdot)$ is non-decreasing
- $\boxed{F(\cdot) \text{ is right-continuous}}$ $\Leftrightarrow F(x) = \lim_{y \downarrow x} F(y)$
- 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):

$F(x) = \mathbb{P}(X \leq x)$

when $x \to \infty$, $\{X \leq x\} \nearrow \Omega$.

By the continuity from below,

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

when $x \to -\infty$, $\{X \leq x\} \searrow \phi$.

By the continuity from above,

$$\lim_{x \to -\infty} \mathbb{P}(X \leq x) = \mathbb{P}(\phi) = 0.$$

UNIVERSITY OF TORONTO

(Proof : $F(x)$ is non-decreasing)

Let $x_1 \le x_2$ . .

Then $F(x_1) = \mathbb{P}(X \le x_1)$

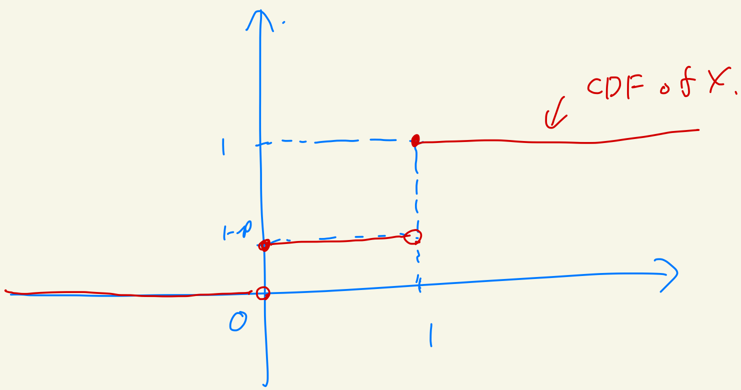
$F(x_2) = \mathbb{P}(X \le x_2)$

Since $x_1 \le x_2$, $\{X \le x_1\} \subset \{X \le x_2\}$.

Therefore, by monotonicity of $\mathbb{P}$, we conclude

$$\mathbb{P}(X \le x_1) \le \mathbb{P}(X \le x_2).$$

example) $\mathbb{P}(X=0) = 1-p$, $\mathbb{P}(X=1) = p$, i.e. $X \sim \text{Bern}(p)$.



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: If X is discrete.

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

sum of mass function

UNIVERSITY OF TORONTO

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

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

continuous case.

density

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:

$X \sim N(\mu, \sigma^2)$, $\quad f(x) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$.

UNIVERSITY OF TORONTO

Independence of random variables

Define independence of random variables based on independence of events:

Independence of random variables

Suppose $X_1, X_2, \cdots, X_n$ are random variables on $(\Omega, \mathcal{F}, P)$, then

$$
\begin{aligned}
& X_1, X_2, \cdots, X_n \text{ are independent} \\
\Leftrightarrow\ & \underline{\{X_1 \in A_1\}, \{X_2 \in A_2\}, \cdots, \{X_n \in A_n\} \text{ are independent,}} \quad \forall A_i \in \mathcal{R} \\
\Leftrightarrow\ & P(\cap_{i=1}^n \{X_i \in A_i\}) = \prod_{i=1}^n P(\{X_i \in A_i\})
\end{aligned}
$$

in terms of events

$\forall A_i \in \mathcal{R}$.

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

$\mathbb{P}(\{X_1 \in A_1\} \cap \{X_2 \in A_2\})$

$= \mathbb{P}(X \in A_1)\,\mathbb{P}(X \in A_2)$

Actually, we need to also check this is true for any $A_i \in \mathbb{R}$.

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

(i) $x_1 < 0$

(ii) $0 \leq x_1 < 1$

(iii) $1 \leq x_1$

Remark:
How to check independence in practice?

Independence of random variables

Corollary of independence

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

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

Choosing $A_i = (-\infty, x_i]$, the definition of independence implies

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

This result says you don't have to check every $A_i \in \mathbb{R}$, you only need to check $A_i = (-\infty, x_i]$

Independence of random variables

Corollary of independence

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

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

Remark:

Independence of discrete random variables

Suppose $X_1, \cdots, X_n$ can only take values from $\{a_1, \cdots\}$, 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?