UNIVERSITY OF TORONTO | Statistical Sciences

DoSS Summer Bootcamp Probability Module 2

Miaoshiqi (Shiki) Liu

University of Toronto

July 13, 2022

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

UNIVERSITY OF TORONTO

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:

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

- 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 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: $A \cap B = B \cap A$ $\quad$ $P(B \cap A) = P(A)P(B) = P(B)P(A)$

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

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

$\{1, 2, 3, \cdots n\}$

$\{1, 2, 3\}$

$\{4, 7, 9, \cdots n\}$

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

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

$\{1, 2, 3\}$

$$P(A_1 \cap A_2 \cap A_3) = P(A_1) P(A_2) P(A_3)$$

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

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

$P(A) = P(\{HH, HT\}) = P(\{HH\}) + P(\{HT\}) = \frac{1}{2}$.

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

Independence of events

Example:

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

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

$P(A)P(B)P(C) = (\frac{1}{2})^3 = \frac{1}{8}$

Independence of events

$P(\cdot \mid C)$ $\tilde{P}(\cdot)$ $P(\quad)$

Conditional independence

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

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

$$\tilde{P}(A \cap B) = \tilde{P}(A)\,\tilde{P}(B)$$

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

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:

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

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

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

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

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

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

Independence of events

Conditional independence

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

$$P(A \cap B \mid C) = P(A \mid C)P(B \mid C). \quad *_1$$

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:

$$\underline{P(A \mid B, C) = P(A \mid C).} \quad *_2$$

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

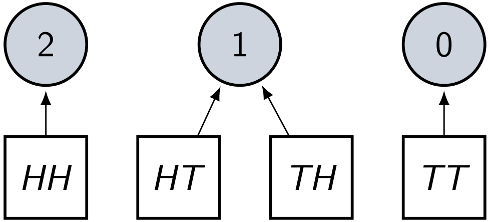


Figure: Mapping from the sample space to the numbers of heads

Random Variables

Example:

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

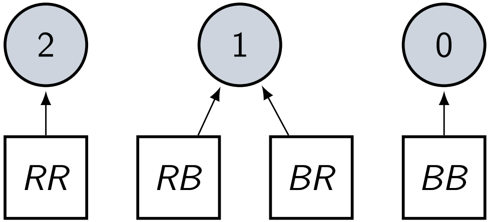


Figure: Mapping from the sample space to the numbers of red balls

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

$\Omega \longrightarrow \mathbb{R}$

$(\Omega, \mathcal{F}, P) \longrightarrow (\mathbb{R}, \mathcal{F}_?, \mu_?)$

$\mathcal{B}(\mathbb{R}) = \mathcal{R}$.

Merits:

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

$P(X=a)$ $\quad P(X \le a)$

Random Variables

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

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

$X(\omega) \in \mathbb{R}$.

$\omega \in \Omega$.

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.

$X \le x$

$X = x$

$X > x$

$X \in (-\infty, x]$

$X \in \{x\}$

$X \in (x, \infty)$

$f(x) = y$.

$y \le 2$

$y = 2$

$f(x) \le 2 \Rightarrow x \in (\cdots)$

$f(x) = 2 \Rightarrow x \in (\cdots)$

UNIVERSITY OF TORONTO

Distribution functions

$\mathbb{R} \quad 2^{\mathbb{R}}$

$X \quad X(\omega)$

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

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

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

$A = (-\infty, x]$

$\mu(A) = \mu((-\infty, x]) = P(X \in (-\infty, x])$

$= P(X \le x)$

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

$\mathbb{R}$.

$\underbrace{(a, b)}$. $a, b \in \mathbb{R}$.

$\bigcup_{n=a}^{\infty} (a, n) \Rightarrow \underline{(a, \infty)}\ \forall a$.

$(-\infty, b), \forall b$

$\underbrace{(-\infty, a], [b, +\infty)}_{a=b,\ \{a\}}$

$X \le a$

$X > a$

$X = a$

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:

$\{1\}$

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

$\mu(\{1\})$

$= P(X=1)$

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

$P(X \leq a)$

$P(X > a) = 1 - P(X \leq a)$

$\{x\}$ $[x, \infty)$ (x, ∞)

$(-\infty, x]$

$P(X > a)$

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:

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

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

$P(X \leq a)$

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

UNIVERSITY OF TORONTO

Distribution functions

Proofs of properties of CDF (first 2 properties):

- $\lim_{x \to \infty} F(x) = 1$

$F(x) = P(X \le x) = \mu((-\infty, x])$

Assume $\{x_n\} \nearrow$ $x_n \le x_{n+1}$ $x_n \to \infty$

$(-\infty, x_n] \subseteq (-\infty, x_{n+1}]$

$\bigcup_{n=1}^{\infty} (-\infty, x_n] = (-\infty, \infty) = \mathbb{R}.$

By continuity from below.

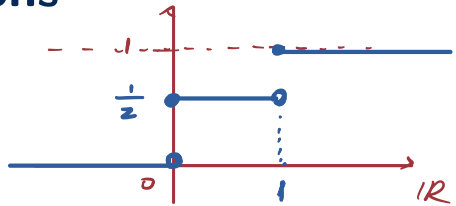
$F(x) \nearrow \mu((-\infty, +\infty)) = 1.$

$\lim_{x \to -\infty} F(x) = 0.$

- $x_1 \le x_2,$ $(-\infty, x_1] \subseteq (-\infty, x_2]$

$F(x_1) = \mu((-\infty, x_1]) \le \mu((-\infty, x_2]) = F(x_2)$

Density functions and mass functions



$X = 1\{H\}$

$X = 0$ or 1

$p(X=0) = \frac{1}{2}$

$= p(X=1)$

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.

$a < 0$

$F(a) = p(X \leq a)$

$= 0$

$0 \leq a < 1$

$F(a) = p(X \leq a)$

$= p(X=0)$

$= \frac{1}{2}$

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:

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

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

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

$$F(x) = \begin{cases} 0, & x < 0 \\ \frac{1}{2} & 0 \leq x < 1 \\ 1 & x \geq 1 \end{cases}$$

UNIVERSITY OF TORONTO

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

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

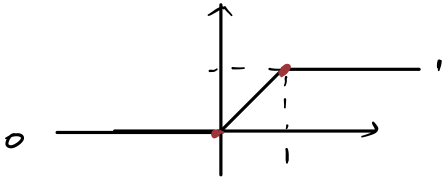
Relationship between PDF and CDF:

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

Example:

$F(x) = x, \quad x \in (0,1)$

$f(x) = \begin{cases} 1, & x \in (0,1) \\ 0, & \text{otherwise.} \end{cases}$



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

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

UNIVERSITY OF TORONTO

Independence of random variables

$X_i = \begin{cases} 0 & \frac{1}{2} \\ 1 & \frac{1}{2} \end{cases}, \quad i = 1, 2.$

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

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

π - λ theorem

π - λ. system.

Corollary of independence

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

$A \in \mathcal{R}$. $A = (-\infty, x]$.

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

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

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

Remark: π - λ theorem.

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

$\{a_i\}$.

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

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?