# Statistical Foundation of Data Science (MA 589)

Lecture Slides

Topic 01: Probability

# Classical Probability

- $S$: Set of all possible outcomes.
- **Definition 1.1:** $P(A) = \frac{\text{Favourable number of cases to } A}{\text{Total number of cases}} = \frac{\#A}{\#S}$.
- **Example 1.1:** A die is rolled. What is the probability of getting 3 on upper face?
  - Ans: $1/6$.
- **Example 1.2:** Consider a target comprising of three concentric circles of radii $1/3$, $1$, and $\sqrt{3}$ feet. What is the probability that a shooter hits inside the inner circle?
  - Both $\#A$ as well as $\#S$ are infinite, the classical probability can not be used here.

# Remarks

- The classical definition works in the first example but does not work in the second.
- Need a better definition which works for wider class of models.
- Start with classical definition and take three key properties to give more general definition of probability.
- Define the probability as a set function.
- Define the domain properly.

# Random Experiment

**Definition 1.2:** An experiment is called a random experiment if it satisfies the following three properties:

1. All the out comes of the experiment is known in advance.
2. The outcome of a particular performance of an experiment is not known in advance.
3. The experiment can be repeated under identical conditions.

**Example 1.3:** Toss a coin.
**Example 1.4:** Toss a coin until the first head appears.
**Example 1.5:** Measuring the height of a student.

# Sample Space

**Definition 1.3:** The collection of all possible outcomes of a random experiment is called the sample space of the random experiment. It will be denoted by $\mathcal{S}$.

**Example 1.6:** $\mathcal{S} = \{H, \ T\}$.
**Example 1.7:** $\mathcal{S} = \{H, \ TH, \ TTH, \ \ldots\}$
**Example 1.8:** $\mathcal{S} = (0, \ \infty)$

# $\sigma$-algebra

**Definition 1.4:** A non-empty collection, $\mathcal{F}$, of subsets of $\mathcal{S}$ is called a $\sigma-$algebra (or $\sigma-$field) if

1. $\mathcal{S} \in \mathcal{F}$
2. $A \in \mathcal{F}$ implies $A^c \in \mathcal{F}$
3. $A_1, A_2, \ldots \in \mathcal{F}$ implies $\cup_{i=1}^{\infty} A_i \in \mathcal{F}$

**Example 1.9:** $\mathcal{F}_1 = \{\phi, \mathcal{S}, \{H\}, \{T\}\}$ ✓, $\mathcal{F}_2 = \{\phi, \mathcal{S}\}$ ✓, $\mathcal{F}_3 = \{\phi, \mathcal{S}, \{H\}\}$ ✗
**Example 1.10:** $\mathcal{F} = \mathcal{P}(\mathcal{S})$ ✓
**Example 1.11:** $\mathcal{F} = \{\phi, \mathcal{S}, (4, 5), (4, 5)^c\}$ ✓

# Events

**Definition 1.5:** A set $E \in \mathcal{F}$ is said to be an event. We will say "the event $E$ occurs" if the outcome of a performance of the random experiment is in $E$.

**Example 1.12:** In measuring height of a student, it turns out to be 4.5 feet. We will say the event $(4, 5)$ has occurred.

# Axiomatic Definition of Probability

**Definition 1.6:** A set function $P : \mathcal{F} \to \mathbb{R}$ is called a probability if

1. $P(E) \geq 0$ for all $E \in \mathcal{F}$
2. $P(\mathcal{S}) = 1$
3. Let $E_1, E_2, \ldots \in \mathcal{F}$ be a sequence of disjoint events then

$$P\left(\cup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} P(E_i)$$

# Examples of Probability

**Example 1.13:** $P(\phi) = 0$, $P(H) = 0.6$, and $P(T) = 0.4$.
**Example 1.14:** For a throw of a die, $\mathcal{S} = \{1, 2, \ldots, 6\}$, $\mathcal{F} = \mathcal{P}(\mathcal{S})$.

1. $P(\phi) = 0$, $P(i) = 1/6$ for $i \in \mathcal{S}$.
2. $P(\phi) = 0$, $P(i) = i/21$ for $i \in \mathcal{S}$.

# Properties of Probability

- $P(\phi) = 0$.
- If $E_1, E_2, \ldots, E_n$ are $n$ disjoint events, then $P\left(\cup_{i=1}^{n} E_i\right) = \sum_{i=1}^{n} P(E_i)$.
- $P$ is monotone, *i.e.*, for $E_1, E_2 \in \mathcal{F}$ and $E_1 \subset E_2$, $P(E_1) \leq P(E_2)$.
- $P$ is subtractive, *i.e.*, for $E_1, E_2 \in \mathcal{F}$ and $E_1 \subset E_2$, $P(E_2 - E_1) = P(E_2) - P(E_1)$.
- $0 \leq P(E) \leq 1$.
- If $E_1, E_2 \in \mathcal{F}$, then $P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2)$.
- If $E_1, E_2 \in \mathcal{F}$, then $P(E_1 \cup E_2) \leq P(E_1) + P(E_2)$.
- If $E \in \mathcal{F}$, then $P(E^c) = 1 - P(E)$.

# Conditional Probability

- A die is thrown twice. What is the probability that the sum is 6?
Ans: $5/36$
- Now suppose you have observed the outcome of the first throw and it is 4. Now what is the probability that the sum will be 6?
Ans: $1/6$.

Once you are given some information or you observe something, the sample space changes. Conditional probability is a probability on the changed sample space.

# Conditional Probability

**Definition 1.7:** Let $H$ be an event with $P(H) > 0$. For any arbitrary event $A$, the conditional probability of $A$ given $H$ is defined by

$$P(A|H) = \frac{P(A \cap H)}{P(H)}.$$

**Remark 1.1:**

$$P(A \cap B) = \begin{cases} P(A)P(B|A) & \text{if } P(A) > 0 \\ P(B)P(A|B) & \text{if } P(B) > 0 \end{cases}$$

# Theorem of Total Probability

**Definition 1.8:** A collection of events $\{E_1, E_2 \ldots\}$ is said to be mutually exclusive if $E_i \cap E_j = \phi, \forall i \neq j$. It is said to be exhaustive if $\cup_i E_i = \mathcal{S}$.

**Theorem 1.1:** Let $\{E_1, E_2 \ldots\}$ be a collection of mutually exclusive and exhaustive events with $P(E_i) > 0, \forall i$. Then for any event $E$,

$$P(E) = \sum_i P(E|E_i)P(E_i).$$

# Bayes Theorem

**Theorem 1.2:** Let $\{E_1, E_2 \dots\}$ be a collection of mutually exclusive and exhaustive events with $P(E_i) > 0, \forall i$. Let $E$ be any event with $P(E) > 0$. Then

$$P(E_i|E) = \frac{P(E|E_i)P(E_i)}{\sum_j P(E|E_j)P(E_j)} \qquad i = 1, 2 \dots.$$

# Bayes Theorem

**Example 1.15:** There are 3 boxes. Box 1 containing 1 white, 4 black balls. Box 2 containing 2 white, 1 black ball. Box 3 containing 3 white, 3 black balls. First you throw a fair die. If the outcomes are 1, 2 or 3 then box 1 is chosen, if the outcome is 4 then box 2 is chosen and if the outcome is 5 or 6 then box 3 is chosen. Finally you draw a ball at random from the chosen box.
a) Given the drawn ball is white what is the (conditional) probability that the ball is from box 1.
b) Given the drawn ball is white what is the (conditional) probability that the ball is from box 2.

- Observe that $P(B_1|W) = 9/34 < 1/2 = P(B_1)$, whereas $P(B_2|W) = 5/17 > 1/6 = P(B_2)$. Thus the "occurrence of one event is making the occurrence of a second event more or less likely".

# Independence

**Definition 1.9:** Let $A$ and $B$ be two events. They are said to be
a) negatively associated if $P(A \cap B) < P(A)P(B)$ ,
b) positively associated if $P(A \cap B) > P(A)P(B)$ ,
c) independent if $P(A \cap B) = P(A)P(B)$ .

- If $P(B) = 0$ then $A$ and $B$ are independent.
- If $P(B) = 1$ then $A$ and $B$ are independent.
- In particular any event $A$ is independent of $\mathcal{S}$ and $\phi$.

**Theorem 1.3:** If $A$ and $B$ are independent, so are $A$ and $B^c$, $A^c$ and $B$, $A^c$ and $B^c$.

# Independence

**Definition 1.10:** A countable collection of events $E_1, E_2, \ldots$ are said to be pairwise independent if $E_i$ and $E_j$ are independent for $i \neq j$.

**Definition 1.11:** A finite collection of events $E_1, E_2, \ldots, E_n$ are said to be independent (or mutually independent) if for any sub-collection $E_{n_1}, \ldots, E_{n_k}$ of $E_1, E_2, \ldots, E_n$,

$$P\left(\bigcap_{i=1}^{k} E_{n_i}\right) = \prod_{i=1}^{k} P(E_{n_i})\,.$$

**Definition 1.12:** A countable collection of events $E_1, E_2, \ldots$ are said to be independent if any finite sub-collection are independent.

# Remarks

- To verify the independence of $E_1, E_2, \dots, E_n$ we must check $2^n - n - 1$ conditions. For example, for $n = 3$, the conditions that need to be checked are $P(E_1 \cap E_2) = P(E_1)P(E_2), P(E_1 \cap E_3) = P(E_1)P(E_3), P(E_2 \cap E_3) = P(E_2)P(E_3), P(E_1 \cap E_2 \cap E_3) = P(E_1)P(E_2)P(E_3)$.
- Independence implies pairwise independence.
- Pairwise independence does not imply independence in general. **Example 1.16:** Let $S = \{HH, HT, TH, TT\}$. Suppose all elementary events are equally likely. Let $E_1 = \{HH, HT\}$, $E_2 = \{HH, TH\}$ and $E_3 = \{HH, TT\}$. Then $E_1, E_2, E_3$ are pairwise independent but not independent.
- $P(E_1 \cap E_2 \cap E_3) = P(E_1)P(E_2)P(E_3)$ is also not sufficient. **Example 1.17:** Let $S = \{(i, j) : i = 1, \dots, 6, j = 1, \dots, 6\}$. Suppose all elementary events are equally likely. Define $E_1 = \{\text{1st roll is } 1, 2 \text{ or } 3\}$, $E_2 = \{\text{1st roll is } 3, 4 \text{ or } 5\}$ and $E_3 = \{\text{Sum of the rolls is } 9\}$.

# Conditional Independent

**Definition 1.13:** Given an event $C$ two events $A$ and $B$ are said to be conditionally independent if $P(A \cap B|C) = P(A|C)P(B|C)$.

**Example 1.18:** A box contains two coins: a fair coin and one fake two-headed coin (i.e., $P(H) = 1$). You choose a coin at random and toss it twice. Define the following events.

$$A = \text{First coin toss results in a } H.$$
$$B = \text{Second coin toss results in a } H.$$
$$C = \text{Coin 1 (regular) has been selected.}$$

Then $A$ and $B$ are conditionally independent given $C$. Are $A$ and $B$ independent?