MATH 5010 –Foundations of Statistical Theory and Probability

❖ Introduction to Probability Theory

Instructor: He Wang
Department of Mathematics
Northeastern University

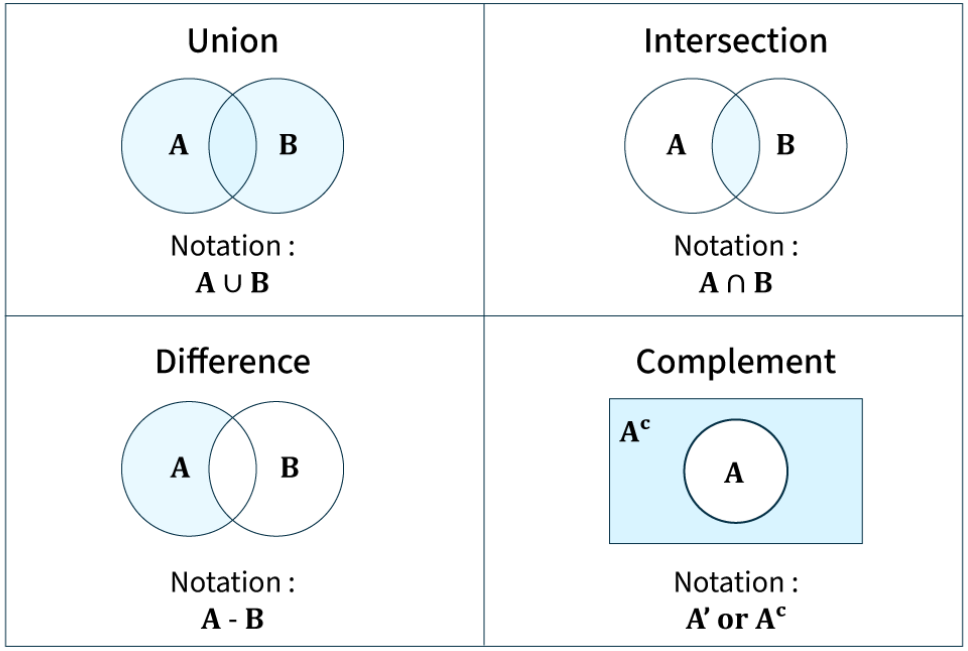
❖ **Outline:**

1. **Set Theory**
2. **Probability Theory**
3. **Combinatorics and Counting**
4. **Conditional Probability and independence**

➢ Set Theory Terminology Review

❑ Standard set **operations**

- "Not A" corresponds to the **complement** $A^c = S \backslash A$;
- "A or B" corresponds to the **union** $A \cup B$;
- "A and B" corresponds to the **intersection** $A \cap B$.



Properties of set operations:

- Commutativity:

$$A \cup B = B \cup A$$
$$A \cap B = B \cap A$$

- Associativity:

$$(A \cup B) \cup C = A \cup (B \cup C)$$
$$(A \cap B) \cap C = A \cap (B \cap C)$$

- Distributive laws:

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$
$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$

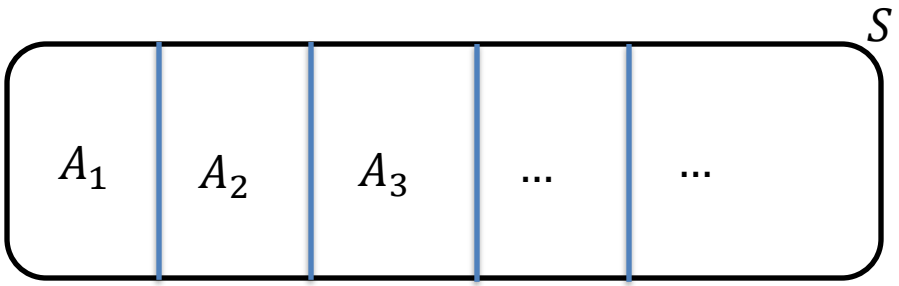
- DeMorgan's Laws:

$$(A \cup B)^c = A^c \cap B^c$$
$$(A \cap B)^c = A^c \cup B^c$$

- Events $A_1, A_2, \ldots$ are **pairwise disjoint/mutually exclusive** if $A_i \cap A_j = \emptyset$ for all $i \neq j$

- A **partition** of S is a sequence of pairwise disjoint sets $A_1, A_2, \ldots$ such that

$$\bigcup_{i=1}^{\infty} A_i = S$$



Example:

$$[0, +\infty) = \bigcup_{i=1}^{\infty} A_i \qquad \text{where } A_i = [i, i+1)$$

- We use $|A|$ to denote the number of elements (**cardinality**) in A.
- A set S is **countable**, if $|S| \leq |\mathbb{N}|$, the cardinality of set of natural numbers.

The set of rational numbers is countable. https://en.wikipedia.org/wiki/Countable_set

➢ Sample Space

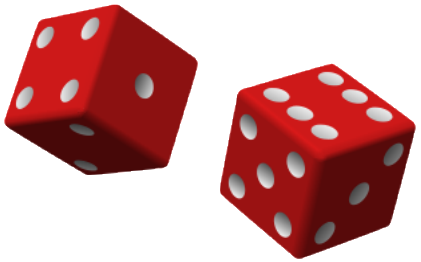
- **Experiment:** A repeatable procedure with a set of possible results (outcomes/realizations/elements).
- **Sample Space S ={all possible outcomes of an experiment}**
- **Event:** A subset of S.

Example 1. Experiment: Flipping a Coin once.

S={Face, Tail}

Example 2. Experiment: Rolling a 6-sided die once.

S={1, 2, 3, 4, 5, 6}

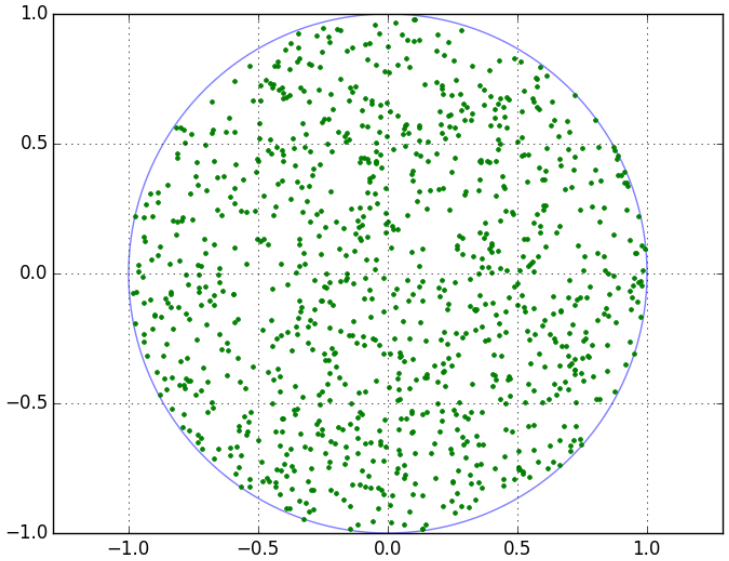


Example 3. Toss a coin repeatedly until Heads appears and record the number of tosses needed:

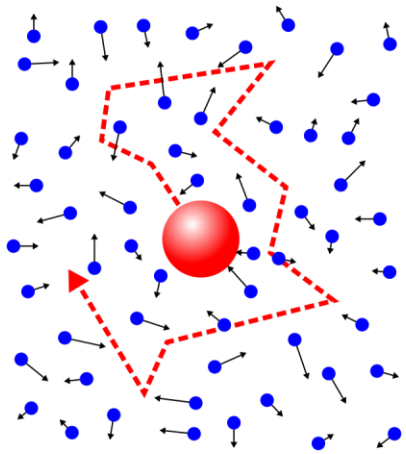
$S = \{1,2,3, \ldots\}$ **infinite**, countable, discrete set

Example 4. Randomly drop a point in a disc.

$$S = \{\vec{x} \in \mathbb{R}^2 \mid x_1^2 + x_2^2 < 1\}$$



Example. (**Brownian motion**) Observe the path of a particle suspended in a liquid or gaseous medium:



$S =$ the set of all continuous paths (functions)

so S is continuous but not finite-dimensional.

Example. Toss a coin infinitely many times:

$S =$all infinite sequences HHHTH…

$S \leftrightarrow [0,1] \subset \mathbb{R}^1$

➢ The Probability Measure (Function)

Definition (*1930s, Kolmogorov*)

Let S be a sample space. Let $\mathcal{B}$ be a collection of subsets of S.

A **probability function/measure** P is a function

$$P: \mathcal{B} \to \mathbb{R}$$

satisfying the following three axioms:

1.) $P(A) \geq 0$ for all $A \in \mathcal{B}$

2.) $P(S) = 1$

3.) $P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$ if $A_i \cap A_j = \emptyset$ for any $i \neq j$

The probability triple $(S, \mathcal{B}, P)$ is called a **probability space.**

Sigma algebra (or Borel Field)

The probability function is not well-defined on any set of subsets of S.

The reason to define the next measurable set is that we want to make sure the probability function is well-defined.

Definition: A **σ-algebra $\mathcal{B}$** is a collection of subsets of S satisfying:

(A1) (full and null set) $S \in \mathcal{B}, \emptyset \in \mathcal{B}$

(A2) (complement) $A \in \mathcal{B} \Rightarrow A^c \in \mathcal{B}$.

(A3) (countably union) $A_1, A_2, \ldots \in \mathcal{B} \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{B}$.

The sets in $\mathcal{B}$ are said to be **measurable** and $(S, \mathcal{B})$ is a **measurable space**.

Note: This course will not work on measure theory.

Some properties from the axioms

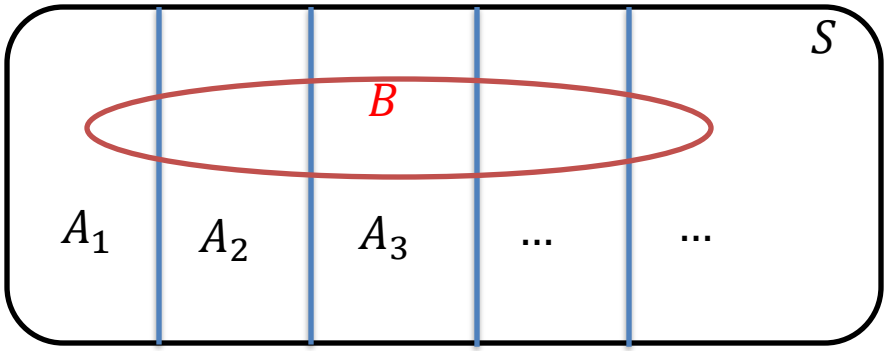
- $P(\emptyset) = 0$
- $0 \leq P(A) \leq 1$
- $A \subset B \Rightarrow P(A) \leq P(B)$
- $P(A^c) = 1 - P(A)$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Example (Bonferroni's Inequality) $P(A \cap B) \geq P(A) + P(B) - 1$

$$P(A \cup B) \leq P(A) + P(B)$$

- **Law of total probability:** For any partition $A_1, A_2, \ldots$ of S,

$$P(B) = P(B \cap A_1) + P(B \cap A_2) + P(B \cap A_3) + \cdots$$



Probability is a continuous Event Function.

We say that a sequence events $A_1, A_2, \ldots$, is an **increasing** sequence if $A_n \subseteq A_{n+1}$ for all $n \geq 1$. Similarly for decreasing sequence.

- If $A_n \subseteq A_{n+1}$, then $P(\bigcup_{n=1}^{\infty} A_n) = \lim_{n\to\infty} P(A_n)$
- If $A_n \supseteq A_{n+1}$, then $P(\bigcap_{n=1}^{\infty} A_n) = \lim_{n\to\infty} P(A_n)$

We define the **limit** as $\lim_{n\to\infty} A_n = \bigcup_{n=1}^{\infty} A_n$ for an an increasing sequence

Similarly, we define the **limit** $\lim_{n\to\infty} A_n = \bigcap_{n=1}^{\infty} A_n$ for a decreasing sequence

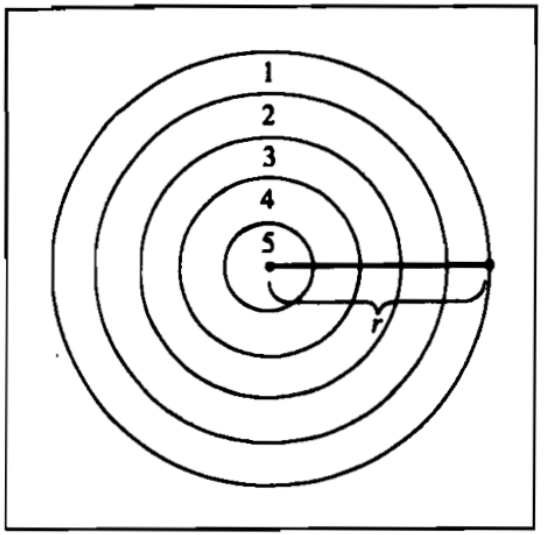
Using the limit notation, the above two formulas for either increasing or decreasing sequences as

$$P\left(\lim_{n\to\infty} A_n\right) = \lim_{n\to\infty} P(A_n)$$

Example. Rolling a 6-sided die once.

Example. flip an **unfair** coin once.

Example: Dart board



The dart board has radius r and the distance between rings is $r/5$. If we make the assumption that the board is always hit, then we have

$$P(\text{scoring } i \text{ points}) = \frac{\text{Area of region i}}{\text{Area of dart board}}$$

$$P(\text{scoring } i \text{ point}) = \frac{(6-i)^2-(5-i)^2}{5^2}$$

1. Classical definition of probability by counting: (a special case for probability function)

Suppose the outcomes of an experiment are <u>all equally likely</u>, and the total number of all possible outcomes is <u>finite.</u>

$$\textit{Probability of an event} := \frac{\textit{Number of ways it can happen}}{\textit{Total number of all possible outcomes}}$$

That is, $$P(A) := \frac{|A|}{|S|}$$

Example 1. Flipping a **fair** coin **2 times**

S={FF, FT, TF, TT}

➢ Combinatorics and Counting

- **Fundamental Theorem of Counting:** If operation A can be performed in m different ways and operation B in n different ways, the sequence (operation A, operation B) can be performed in $m \times n$ different ways.

- **Permutation rule**: The number of ways to arrange k objects of a set of n distinct elements (permutations)

$$P_k^n = \frac{n!}{(n-k)!} = n(n-1)\cdots(n-k+1)$$

- **Combination Rule:** The number of ways to choose a subset of k objects from n distinct objects (combinations)

$$C_k^n := \binom{n}{k} = \frac{n!}{(n-k)!\,k!} = \frac{P_k^n}{k!}$$

Four Possible Methods of Counting

1. Replacement = Yes, Order = Yes.

- You're choosing r objects from n options, and you can repeat choices.
- **Example:** 3-digit codes using digits 0–9:
- Number of ways **(**Product Rule**)** $n^r = 10^3 = 1000$.

2. Replacement = No, Order = Yes.

- You're choosing r objects from n without replacement, but the sequence matters.
- **Example:** 3-letter arrangements from 5 letters (e.g., ABCDE)
- Number of ways **(Permutations)**

$$P_3^5 = \frac{5!}{(5-3)!}$$

3. Replacement = No, Order = No:

- Choosing r distinct objects from n, and the order doesn't matter.
- **Example:** Choosing 3 students from 10 for a committee
- Number of ways **(Combination)**

$$\binom{n}{r} = \binom{10}{3} = 120$$

4. Replacement = Yes, Order = No.

- You choose r objects from n, and repetition is allowed, but order doesn't matter.
- **Example:** Choosing 3 scoops of ice cream (flavors can repeat) from 5 flavors.
- Number of ways: $\binom{n+r-1}{r} = \binom{7}{3} = 35$

Better to think put 3 scoops of ice cream to 5 bins.

2. Frequency definition

The probability of an event is the limit of the relative frequency with which that event occurs in a large number of repeated trials under identical conditions.

$$\textit{Probability of an event} := \frac{\text{Times an event happens}}{\textit{Total number} \text{ experiment is repeated } \textbf{many} \text{ time}}$$

Coin Toss Example: If you flip a coin many times, the probability of getting heads is considered to be 0.5 because in the long run, the number of heads will be approximately half the total number of flips.

➢ Conditional Probability

Definition. Probability that event A occurs **given** that event B already occurs, denoted by **P(A|B),** is a **conditional probability of A given B**, defined by

$$P(A|B) := \frac{P(A \cap B)}{P(B)} \qquad \text{if } P(B) \neq 0$$

When B is fixed, the function $P(\cdot\,|B) : S \to \mathbb{R}$ is another probability measure.

Example. Rolling a fair 6-sided die once.

Given that the result is an even number, what is the probability that the result is 6?

➢ Independence

Definition: The **events** $A \; and \; B$ are called **independent** if

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

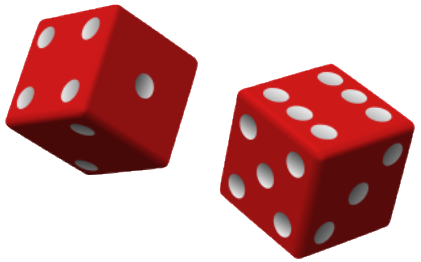
If $A \; and \; B$ are not empty set, A and B are **independent** if and only if

$$P(A|B) \; = \; P(A) \text{ if and only if } P(B|A) \; = \; P(B)$$

That is, knowing B does NOT change the probability of A. The same for the other way.

Example. Rolling a 6-sided die twice.

A: the first face is even number
B: the second face is 6.



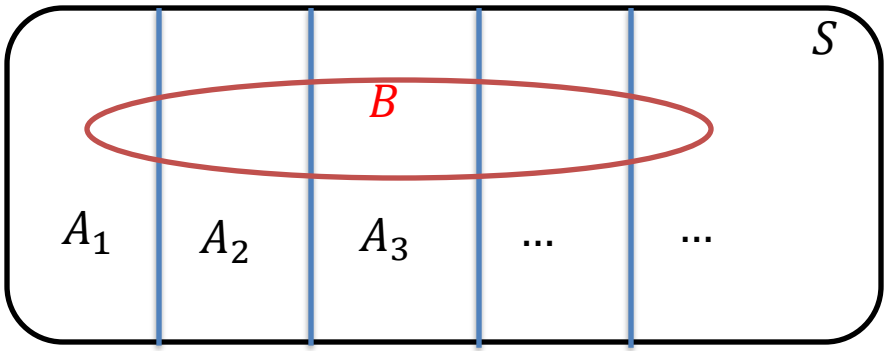
Theorem 1. Law of Total Probability

Let $A_1, A_2, \ldots, A_n$ be a sequence of events in a sample space S such that

$$S = \bigcup_{i=1}^{n} A_i \qquad \text{and} \qquad A_i \cap A_j = \emptyset \text{ for all } i \neq j$$

Then, for any event B,

$$P(B) = \sum_{i=1}^{n} P(B \cap A_i) = \sum_{i=1}^{n} P(B|A_i)\, P(A_i)$$



Theorem 2. Bayes' Theorem

Let $A_1, A_2, \ldots, A_n$ be a sequence of events in a sample space S such that

$$S = \bigcup_{i=1}^{n} A_i \qquad\qquad A_i \cap A_j = \emptyset \text{ for all } i \neq j$$

Then, for any event B,

$$P(A_j|B) = \frac{P(B, A)}{P(A)} = \frac{P(B|A_j)P(A_j)}{\sum_{i=1}^{n} P(B|A_i)P(A_i)}$$

Some examples of Bayes' rule application

1. Covid-19 test.
2. Monty Hall Problem.
3. Bayesian inference example

Example: Testing for Covid-19.

$D = 1$ Infected by disease. ($D = 0$ not infected.)

$Y = 1$ **Test** positive. ($\mathrm{Y} = 0$. Test negative) –binary classification for Y

Test **Sensitivity (True-Positive Rate)**: $= P(Y = 1|D = 1)$

Test **Specificity (True-Negative Rate)**:$= P(Y = 0|D = 0)$

Prevalence of the disease $= P(D = 1)$

Suppose $P(D = 1) = 0.01$.

$P(Y = 1|D = 1) = 0.875$

$P(Y = 0|D = 0) = 0.975$

Then suppose a person test positive, what is the chance that the person really infected?

$$P(D=1|T=1) = \frac{P(Y=1|D=1)P(D=1)}{P(Y=1|D=1)P(D=1) + P(Y=1|D=0)P(D=0)}$$

$$= \frac{0.875 * 0.01}{0.875 * 0.01 + 0.025 * 0.99}$$

$$= 0.26$$

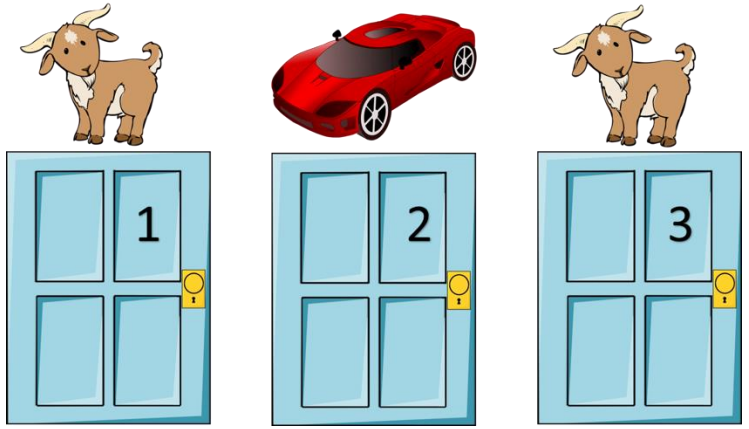
Similarly, we also know the chance that a person infected given test negative:

$$P(D=1|T=0) = \frac{P(Y=0|D=1)P(D=1)}{P(Y=0|D=1)P(D=1) + P(Y=0|D=0)P(D=0)}$$

$$= 0.0013$$

Monty Hall Problem (https://en.wikipedia.org/wiki/Monty_Hall_problem)

In a game show, there are three doors with is a big prize behind only one door. You choose one of them, then one of the left is opened and there is no prize behind the opened door. You have a chance to switch your choice. Will you switch?



C: 1st choice is correct. $P(C) = 1/3$
S: win after switch.
T: win without switch.

$$P(S) = P(S|C)P(C) + P(S|C^c)P(C^c) = 0(\frac{1}{3}) + 1(\frac{2}{3}) = \frac{2}{3}$$

$$P(T) = P(T|C)P(C) + P(T|C^c)P(C^c) = 1(\frac{1}{3}) + 0(\frac{2}{3}) = \frac{1}{3}$$

Conclusion: you should switch.

Bayesian Inference

I tell you that I can toss coin such that it always comes up Heads. You are 95% certain that I am lying. I tossed a coin 5 times in front of you and comes up Head every time. How certain are you now that I am lying

H_1: I am lying (the coin if fair).
$H_2 = H_1^c$: I can always toss Heads.
D=Data:= { I tossed 5 times and got 5 heads }
Prior probabilities: $P(H_1) = 0.95$ and $P(H_2) = 0.05$
Posterior probabilities: (after experiments) $p_1 = P(H_1|D)$ updated probability for H_1 given data D. By Bayes's Theorem,

$$P(H_1|D) = \frac{P(D|H_1)P(H_1)}{P(D|H_1)P(H_1) + P(D|H_2)P(H_2)} == \frac{(0.5)^5 0.95}{(0.5)^5 0.95 + 1(0.05)} = 0.37$$

Summary: based on the data D, your new degree of certainty that I am lying is 37%.
We used independence to calculate $P(D|H_1) = P(HHHHH) = P(H)^5 = 0.5^5$.

Further examples:
1. Naive Bayes spam filtering https://en.wikipedia.org/wiki/Naive_Bayes_spam_filtering
2. Bayesian poisoning https://en.wikipedia.org/wiki/Bayesian_poisoning

❖ Conditional Independence

- **Independence**

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

- **Conditional Independence**

  Two events $A\ and\ B$ are **conditional independent given $\boldsymbol{C}$**, if

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

  Two Random Variables $X\ and\ Y$ are **conditional independent given $\boldsymbol{Z}$**

$$F_{X,Y|Z}(x, y|z) = F_{X|Z}(x|z)F_{Y|Z}(y|z) \text{ for all } x, y, z$$

Remark: there is no direct relation between independence and conditional independence. (i.e., one can not implies the other one.)

➢ History of Probability

- The concept of probability can be traced back to ancient civilizations and Greek philosophy.
- The modern study of probability began in the 16th century, e.g., “Book on Games of Chance” by G. Cardano -1564
- Ground work of Probability theory by B. Pascal and P. Fermat in 1650’s.
- J. Bernoulli (1713) and P. Laplace (1812) further developed the theory, including law of large numbers, Bayesian probability and establishing the foundations of statistical inference .
- Kolmogorov’s work in the formalization of probability theory in early 20^{th} century. Further development includes measure-theoretic definitions, Martingales and Stochastic Processes, etc.
- The formalization of statistical methods in the early 20th century including R. Fisher who developed techniques for experimental design and analysis.
- Computational Probability including Monte Carlo simulations and applications.

Frequentist and Bayesian (first meet)

- Frequentist view: the probability is interpreted as the limiting frequencies observed over repetitions in identical situations.
- Bayesian/subjective view: where the probability quantifies personal belief.

The probability of an event E is the price one is just willing to pay to enter a game in which one can win a unit amount of money if E is true.

Example: If I believe a coin is fair and am to win 1 unit if a head arises, then I would pay 1/2 unit of money to enter the bet.

References:

- **Book 1. [CB] Statistical Inference**, by Casella, George, Berger, Roger L, 2nd edition
- **Book 2. [W]: All of Statistics: Larry Wasserman**
- **Book 3. Introduction to Probability**. C.M. Grinstead and J.L. Snell. American Mathematical Society, 2012
- **Book 4. Introduction to Probability Models**, S. Ross, 12th edition (published by Academic Press).

Online books:

- https://www.probabilitycourse.com/
- https://online.stat.psu.edu/stat415/
- https://stat110.hsites.harvard.edu/
- https://bookdown.org/egarpor/inference/

Extra Reading:

Baby Measure Theory: https://www.stat.umn.edu/geyer/8501/measure.pdf

YouTube video about coin flips by a famous statistician,
YouTube video about dice rolls (Part I and Part II).