1 Welcome to 446.358

1.1 What is this course about?

446.358 is a course on mathematical probability theory with

- Applications to the statistical problems of decision-making under uncertainty and estimation of parameters
- Applications to problems in engineering

The topics to be covered will provide essential preparation for many senior-level and graduate courses in communication systems, signal processing, filtering, computer networks, reliability of systems, stochastic controls. After this course, you will be able to apply probabilistic reasoning to understand how and why various statistical procedures work.

1.2 What is probability?

Since pre-historic times, Mankind has noticed Random phenomena such as coin tosses, rolling dice, horse races. The Babylonians, Chinese, Egyptians, Greeks, Persians, Romans, Vikings, the Aryans in ancient India, all gambled, and Aristotle discussed dice probabilities.

Modern mankind also loves to gamble. Nowadays, governments even encourage gambling through State-run lotteries, Casinos, Betting Parlors, etc.

Probabilistic notions are commonplace in everyday language usage. We use words such as probable/improbable; possible/impossible; certain/uncertain; likely/unlikely.

1.3 Axiomatic Approach to Probability

In the axiomatic approach, probabilities are numbers in the range [0,1]. Certain probabilities are assumed to be given (we don't ask how!). The axiomatic approach allows the calculation of other probabilities in a mathematically and logically consistent manner.

We will build a mathematical model of probabilities that will allow us to draw useful and interesting conclusions about the real world. The model does not describe all aspects of probability but enough aspects are captured to make the model very useful.

1.4 Probability versus Statistics

Statistics are collections of data. Statistics is the science of collecting data, organizing data, analyzing data, and making inferences from data. In this course, we shall study a few selected topics in statistical inference.

Probability theory deals with questions such as What is the probability that Heads shows six times on ten tosses of a fair coin? Whether the coin actually is fair is not open to question - fairness is one of the given in the problem. The probability question can be answered precisely and unambiguously.

Probability versus Statistics l Statistical inference deals with questions such as "Given that Heads showed on six of ten tosses, is it reasonable to assume the coin is fair?" A definitive answer is not possible since different people have different ideas as to what is reasonable. Statistical inferences are accompanied by levels of confidence, such as "Yes, I am 95% sure the coin is fair." NO statistical procedure can prove or disprove that the coin is fair. Fairness is a matter of belief, and the degree of belief is the confidence level.

1.5 Probability at Work

1.5.1 Probability in Physics

• Maxwell-Boltzmann kinetic theory of gases: Motion of gas molecules is random. Gas laws can be deduced from the aggregate behavior of many molecules

• Statistical mechanics

• Quantum mechanics: At the atomic level, physical phenomena can only be described probabilistically

• Semiconductor physics and electronics: Many practical devices in use these days are designed to make use quantum-mechanical effects

1.5.2 Probability in Engineering

• Thermal noise in electrical circuits

• Detection of weak radio and radar signals

• Information theory

• Communication systems design

• Reliability of systems : Failure probabilities, Failure rates, Mean time to failure

• Networks and Systems Problems :Random arrivals of packets/jobs, Random lengths/service times, Random requests for resources, Probability of buffer or queue overflow, Transmission or service delays, Scheduling problems, priorities, QOS, Flow control and routing

In this course, we emphasize the basic science and math. You need to take follow-on courses to gain better understanding of where and how probability is used in engineering.

1.6 Things you know already on counting

1.6.1 permutation

Suppose that we have n different objects. How many different ordered arrangements (=permutations) are possible?
Ans. $n(n-1)(n-2)\cdots(n-r+1) \ = n!$ different permutations

1.6.2 combination

How many different groups of r objects could be formed from a total of n objects?
Ans. the number of possible combinations of n objects taken r at a time= $\frac{n(n-1)\cdots 2\cdot 1}{r!} = \frac{n!}{(n-r)!r!} = \left(\begin{array}{c} n \\ r \end{array} \right)$

1.6.3 binomial theorem

$$(x+y)^n = \sum_{k=1}^{n} \left(\begin{array}{c} n \\ r \end{array} \right) x^k y^{n-k}$$

1.6.4 multinomial theorem

When $n_1 + \cdots + n_r = n$, the number of possible divisions of n distinct objects into r distinct groups of respective sizes $n_1, \cdots, n_r$ is

$$\left(\begin{array}{c} n \\ n_1, \cdots, n_r \end{array} \right) = \frac{n!}{n_1! \cdots n_r!}$$

and

$$(x_1 + \cdots + x_r)^n = \sum_{(n_1,\cdots,n_r):n_1+\cdots+n_r=n}^{n} \left(\begin{array}{c} n \\ n_1, \cdots, n_r \end{array} \right) x_1^{n_1} \cdots x_n^{n_r}.$$

The sum is over all nonnegative integer-valued vectors $(n_1, \cdots, n_r)$ such that $n_1 + \cdots + n_r = n$.

1.7 Axioms of Probability

1.7.1 Sample Space

- An experiment is performed and its outcome observed. This is called a trial of the experiment.
- The set of all possible outcomes of an experiment is called the sample space Ω of the experiment

Example . The experiment is tossing a coin: $\Omega = \{H, T\}$

Example . The experiment is rolling a die: $\Omega = \{1, 2, 3, 4, 5, 6\}$

For rolling a dice, suppose that each outcome is equally likely: $P(1) = P(2) = \cdots = P(6) = 1/6$. What is the probability of rolling an even number? What is the probability of rolling a prime number? P(even number) = P(prime number)=1/2 also (1 is not a prime).

1.8 Events

A subset of Ω is called an event.
Example . $A = \{2, 4, 6\}$ and $B = \{2, 3, 5\}$ are said to be events defined on the sample space

$\Omega = \{1, 2, 3, 4, 5, 6\}$
Example . $A^c = \{1, 3, 5\}$ and $B^c = \{1, 4, 6\}$ also are events defined on Ω.

• "events defined on the sample space" is merely a probabilists way of saying "subsets of the sample space".
• An event A is said to have occurred on a trial if the outcome of the trial is a member of the subset A (we dont care which member of A it is).
• If the observed outcome is not a member of A, then we say A did not occur, or equivalently, we say that A^c occurred
• Outcomes versus Events

. Every trial results in only one outcome, that is, only one of the elements in W can be the observed outcome
. The observed outcome is a member of several different subsets, i.e., events, and all these events are said to have occurred
. Fundamental notion: On each trial of the experiment, one outcome occurs, but many events occur

Example . If the outcome of rolling a die is 4, then

Events $A = \{2, 4, 6\}$ and $B^c = \{1, 4, 6\}$ both have occurred.
Events $A^c = \{1, 3, 5\}$ and $B = \{2, 3, 5\}$ did not occur.
Event $A \cup B^c = \{1, 2, 4, 6\}$ has occurred
Event $A \cap B^c = \{4, 6\}$ also has occurred

• For any event A, exactly one of the two events $A$and A^c occurred, and the other did not.
• On any trial, the event Ω always occurs. The event Ω is called the certain event or the sure event.
• On any trial, the event $\emptyset$ never occurs. The event $\emptyset$ is called the null event or the impossible event.
• Regardless of how we choose to assign probabilities to the outcomes, $P(\Omega) = 1, P(\emptyset) = 0$
• How many events are there in all?

. A sample space Ω of n elements has 2^n different subsets including Ω and $\emptyset$. $2^n - 1$ of these subsets are nonempty.
. The 2^n events can be paired up into 2^{n-1} pairs of the form $\{A, A^c\}$. $\{\Omega, \emptyset\}$ is one such pair.

. On each trial of the experiment, exactly 2^{n-1} events (one from each pair) occur and 2^{n-1} events (other from each pair) do not.

• Events A and B are said to be disjoint or mutually exclusive if $A \cap B = \emptyset$. For a disjoint union of events A and B, $P(A \cup B) = P(A) + P(B)$.

Example .

Let $A = \{x_2, x_4, x_{22}\} = \{x_2\} \cup \{x_4\} \cup \{x_{22}\}$. Then $P(A) = P\{x_2\} + P\{x_4\} + P\{x_{22}\} = p_2 + p_4 + p_{22}$

1.9 Probability Axioms for finite spaces

The ideas described thus far are the basis of the axioms of probability theory. Probabilities are numbers assigned to events that satisfy the following rules.

- **Axiom I**: $0 \leq P(A) \leq 1$ for all events A
- **Axiom II**: $P(\Omega) = 1$
- **Axiom III**: If events A and B are disjoint, then $P(A \cup B) = P(A) + P(B)$

Consequences of the Axioms

• Ω and $\emptyset$ are disjoint events.

$$P(\Omega \cup \emptyset) = P(\Omega) + P(\emptyset) = 1 + P(\emptyset)$$

But $\Omega \cup \emptyset = \Omega$.
Hence $P(\Omega \cup \emptyset) = P(\Omega) \implies P(\emptyset) = 0$.
• A and A^c are disjoint events; $A \cup A^c = \Omega$.

$$1 = P(\Omega) = P(A \cup A^c) = P(A) + P(A^c)$$
$$P(A^c) = 1 - P(A); \quad P(A) = 1 - P(A^c)$$

Since $P(A) \geq 0$ and $P(A^c) \geq 0$ (Axiom I), we deduce that

$$0 \leq P(A) \leq 1, 0 \leq P(A^c) \leq 1 \quad \text{for all events } A$$

1.10 Partitions

Sets $\{A, B, C, \cdots, G\}$ are a partition of set H if H is the disjoint union of $A, B, \cdots, G$. Then Axiom III straightforwardly generalizes to give

$$P(H) = P(A) + P(B) + \cdots + P(G) \ .$$

Venn diagram vs. Karnaugh maps

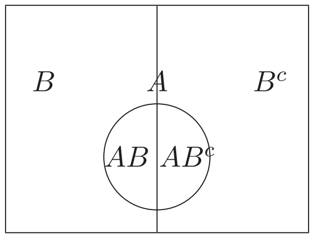

Figure 1: Venn diagram

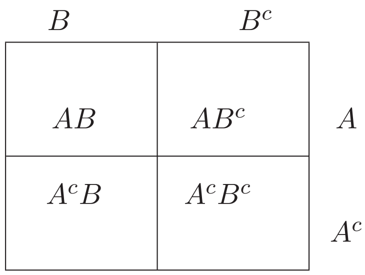

Figure 2: Karnaugh maps

Example .

$A \cap B$ and $A \cap B^c$ are a partition of A.
$A \cap B$ and $A^c \cap B$ are a partition of B.
($A \cap B$ is often abbreviated to just AB.)
AB, A^cB, AB^c, and A^cB^c are a partition of Ω.
$ABC, ABC^c, AB^cC, AB^cC^c, A^cBC, A^cBC^c, A^cB^cC$, and $A^cB^cC^c$ are a partition of Ω.

Q. What is $P(A \cup B)$ when A and B are not disjoint events?
Ans. Partition $A \cup B$ into AB, AB^c, A^cB. Then
P(A $\cup$ B) = P(AB) + P(AB^c) + P(A^{c}B)
= P(AB) + P(AB^c) + P(A^{c}B) + P(AB) - P(AB)
= P(A) + P(B)- P(AB)

Other formulas

- If A and B are disjoint, $P(AB) = P(\emptyset) = 0$.
- $P(A \cup B) = P(A) + P(A^cB) = P(B) + P(B^cA)$
- $P(A \oplus B) = P\{\text{only one of } A \text{ and } B \text{ occur}\} = P(A) + P(B) - 2P(AB)$
- $P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(AB) - P(BC) - P(AC) + P(ABC)$

DeMorgan's Laws

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

1.11 Sample spaces with equally likely outcomes

If $S = \{1, 2, \cdots, N\}$ and all outcomes are equally likely to occur, then

$$P(\{1\}) = P(\{2\}) = \cdots = P(\{N\}) = 1/N \ .$$

Example A.n urn contains 6 identical red and 3 identical green balls. A trial of the experiment consists of simultaneously drawing two balls at random from the urn. The outcomes of this experiment are the subsets of size 2 from the set of 9 balls ($9 - choose - 2 = 36$ subsets). Probability that one of the drawn balls is red and the other is green is

$$\frac{(6 - choose - 1) \times (3 - choose - 1)}{(9 - choose - 2)}$$

1.12 Infinite sample spaces

Example . The experiment consists of tossing a coin till a Tail appears for the first time.

The sample space is $\Omega = \{T, HT, HHT, HHHT, \cdots\}$.

This sample space is countably infinite. **Countable** means there is a one-to-one correspondence between the integers and the outcomes.

$$\text{Integer } n \longleftrightarrow \overbrace{HH\cdots H}^{n-1 \text{ times}}T$$

New improved **Axiom III**: Let $A_1, A_2, \cdots A_n, \cdots$ denote a countable sequence of disjoint events, that is, $A_iA_j = \emptyset$ for all $i \neq j$. Then,

$$P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i) \ .$$

If every outcome is an event of probability zero, then isn't it true that any event A must also have probability zero?

P(A) = sum of the probabilities of all the outcomes that comprise A

= 0 + 0 + ... = 0 ?

No, the above is a mis-application of Axiom III (which applies to countable unions only).

Example . Choose a random number between 0 and 1. Each outcome (and also any countable set of outcomes) has probability zero. However,

$$P\{a < \text{outcome} < b\} = b - a \qquad \text{for } 0 \leq a < b \leq 1 \ .$$

The nonzero probabilities are assigned to the intervals of the line, not to outcomes!