MC4010 - Discrete Mathematics

Discrete Probability

Dr T. Mayooran

Department of Inter Disciplinary Studies, Faculty of Engineering, University of Jaffna

2023-11-15

All contents of these slides are agglomerated from recommended reference textbook of this course.

Slides are prepared through RMarkdown Beamer presentation option and executed via RStudio.

Introduction

- The theory of probability was first developed more than 300 years ago, when certain gambling games were analyzed. Although probability theory was originally invented to study gambling, it now plays an essential role in a wide variety of disciplines.
- For example, probability theory is extensively applied in the study of genetics, where it can be used to help understand the inheritance of traits. Of course, probability still remains an extremely popular part of mathematics because of its applicability to gambling, which continues to be an extremely popular human endeavor.
- In computer engineering, probability theory plays an important role in the study of the complexity of algorithms. In particular, ideas and techniques from probability theory are used to determine the average-case complexity of algorithms.

- Probability theory can help us answer questions that involve uncertainty, such as determining whether we should reject an incoming mail message as spam based on the words that appear in the message.
- Probability theory dates back to 1526 when the Italian mathematician, physician, and gambler Girolamo Cardano wrote the first known systematic treatment of the subject in his book Liber de Ludo Aleae (Book on Games of Chance).
- In the seventeenth century the French mathematician Blaise Pascal determined the odds of winning some popular bets based on the outcome when a pair of dice is repeatedly rolled.
- In the eighteenth century, the French mathematician Laplace, who also studied gambling, defined the probability of an event as the number of successful outcomes divided by the number of possible outcomes.

Finite Probability

An experiment is a procedure that yields one of a given set of possible outcomes. The sample space of the experiment is the set of possible outcomes. An event is a subset of the sample space. Laplace's definition of the probability of an event with finitely many possible outcomes will now be stated.

Definition - 01 If S is a finite non-empty sample space of equally likely outcomes, and E is an event, that is, a subset of S, then the probability of E is

$$P(E) = \frac{n(E)}{n(S)}$$

According to Laplace's definition, the probability of an event is between 0 and 1. To see this, note that if E is an event from a finite sample space S, then $0 \leq n(E) \leq n(S)$, because $E \subseteq S$. Thus, $0 \leq p(E) = \frac{n(E)}{n(S)} \leq 1$.

Examples

1. An urn contains four blue balls and five red balls. What is the probability that a ball chosen at random from the urn is blue?
2. What is the probability that when two dice are rolled, the sum of the numbers on the two dice is 7?
3. In a lottery, players win a large prize when they pick four digits that match, in the correct order, four digits selected by a random mechanical process.A smaller prize is won if only three digits are matched. What is the probability that a player wins the large prize? What is the probability that a player wins the small prize?

Examples

1. There are many lotteries now that award enormous prizes to people who correctly choose a set of six numbers out of the first n positive integers, where n is usually between 30 and 60. What is the probability that a person picks the correct six numbers out of 40?
2. What is the probability that the numbers 11, 4, 17, 39, and 23 are drawn in that order from a bin containing 50 balls labeled with the numbers 1, 2, . . . , 50 if (a) the ball selected is not returned to the bin before the next ball is selected and (b) the ball selected is returned to the bin before the next ball is selected?

Probabilities of Complements and Unions of Events

We can use counting techniques to find the probability of events derived from other events.

Theorem - 01

Let E be an event in a sample space S. The probability of the event $\bar{E} = S - E$, the complementary event of E, is given by

$$P(\bar{E}) = 1 - P(E)$$

There is an alternative strategy for finding the probability of an event when a direct approach does not work well. Instead of determining the probability of the event, the probability of its complement can be found.

Example:

A sequence of 10 bits is randomly generated. What is the probability that at least one of these bits is 0?

Solution: Let E be the event that at least one of the 10 bits is 0. Then E is the event that all the bits are 1s. Because the sample space S is the set of all bit strings of length 10, it follows that

$$P(\bar{E}) = 1 - P(E) = 1 - \frac{n(E)}{n(S)} = 1 - \frac{1}{2^{10}}$$
$$= 1 - \frac{1}{1024} = \frac{1023}{1024}$$

Hence, the probability that the bit string will contain at least one 0 bit is $1023/1024$. It is quite difficult to find this probability directly without using Theorem 1.

Theorem - 02

Let E_1 and E_2 be events in the sample space S. Then

$$P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2)$$

Example:

What is the probability that a positive integer selected at random from the set of positive integers not exceeding 100 is divisible by either 2 or 5?

Solution: Let E_1 be the event that the integer selected at random is divisible by 2, and let E_2 be the event that it is divisible by 5. Then $E_1 \cup E_2$ is the event that it is divisible by either 2 or 5. Also, $E_1 \cap E_2$ is the event that it is divisible by both 2 and 5, or equivalently, that it is divisible by 10.

Because $n(E_1) = 50, n(E_2) = 20$, and $n(E_1 \cap E_2) = 10$, it follows that

$$P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2)$$

$$= \frac{n(E_1)}{n(S)} + \frac{n(E_2)}{n(S)} - \frac{n(E_1 \cap E_2)}{n(S)}$$

$$= \frac{50}{100} + \frac{20}{100} - \frac{10}{100} = \frac{3}{5} = 0.6$$

Practice problem

Proof Theorems 1 and 2 by using your set theory knowledge.

Probabilistic vs Statistical Reasoning

- Suppose we know exactly the proportions of Senaro GN 125 motorbike assembles in Sri Lanka. Then we can find the probability that the first bike we see in the street is a Senaro GN 125. This is probabilistic reasoning as we know the population and predict the sample.
- Now, suppose that we do not know the proportions of Senaro GN 125 motorbike assembles in Sri Lanka, but would like to estimate them. we observe a random sample of Senaro GN 125 bikes in the street and then we have an estimate of the proportions of the population. This is statistical reasoning.

Assigning Probabilities

Let S be the sample space of an experiment with a finite or countable number of outcomes.We assign a probability $p(s)$ to each outcome s.We require that two conditions be met:

condition (1) : $0 \leq P(s) \leq 1$ for each $s \in S$

condition (2) : $\sum_{s \in S} P(s) = 1$

Condition (1) states that the probability of each outcome is a nonnegative real number no greater than 1. Condition (2) states that the sum of the probabilities of all possible outcomes should be 1; that is, when we do the experiment, it is a certainty that one of these outcomes occurs.

The function P from the set of all outcomes of the sample space S is called a probability distribution.

Examples

- What probabilities should we assign to the outcomes H (heads) and T (tails) when a fair coin is flipped? What probabilities should be assigned to these outcomes when the coin is biased so that heads comes up twice as often as tails?

Definition

The probability of the event E is the sum of the probabilities of the outcomes in E. That is,

$$P(E) = \sum_{s \in E} P(s)$$

(Note that when E is an infinite set, $\sum_{s \in E} P(s)$ is a convergent infinite series.)

Example:

Suppose that a die is biased (or loaded) so that 3 appears twice as often as each other number but that the other five outcomes are equally likely. What is the probability that an odd number appears when we roll this die?

Theorem

If $E_1, E_2, \cdots$ is a sequence of pairwise disjoint events in a sample space S, then

$$P\left(\bigcup_i E_i\right) = \sum_i P(E_i)$$

(Note that this theorem applies when the sequence $E_1, E_2, \cdots$ consists of a finite number or a countably infinite number of pairwise disjoint events.)

Conditional Probability

In general, to find the conditional probability of E given F, we use F as the sample space. For an outcome from E to occur, this outcome must also belong to $E \cap F$.With this motivation, we make following definition,

Theorem

Let E and F be events with $P(F) > 0$. The conditional probability of E given F, denoted by $P(E|F)$, is defined as

$$P(E|F) = \frac{P(E \cap F)}{P(F)}$$

Examples

1. A bit string of length four is generated at random so that each of the 16 bit strings of length four is equally likely. What is the probability that it contains at least two consecutive 0s, given that its first bit is a 0? (We assume that 0 bits and 1 bits are equally likely.)
2. What is the conditional probability that a family with two children has two boys, given they have at least one boy? Assume that each of the possibilities BB, BG, GB, and GG is equally likely, where B represents a boy and G represents a girl. (Note that BG represents a family with an older boy and a younger girl while GB represents a family with an older girl and a younger boy.)

Independence

Definition

The events E and F are independent if and only if $P(E \cap F) = P(E)P(F)$.

Example

1. Suppose E is the event that a randomly generated bit string of length four begins with a 1 and F is the event that this bit string contains an even number of 1s. Are E and F independent, if the 16 bit strings of length four are equally likely?

2. Assume, as in Example 2 of slide number 19, that each of the four ways a family can have two children is equally likely. Are the events E, that a family with two children has two boys, and F, that a family with two children has at least one boy, independent?

Solutions

There are eight bit strings of length four that begin with a one: 1000, 1001, 1010,1011, 1100, 1101, 1110, and 1111. There are also eight bit strings of length four that contain an even number of ones: 0000, 0011, 0101, 0110, 1001, 1010, 1100, 1111. Because there are 16 bit strings of length four, it follows that

$$p(E) = p(F) = 8/16 = 1/2.$$

Because $E \cap F = 1111, 1100, 1010, 1001$, we see that

$$p(E \cap F) = 4/16 = 1/4.$$

Because

$$p(E \cap F) = 1/4 = (1/2)(1/2) = p(E)p(F),$$

we conclude that E and F are independent.

Solutions

Because $E = BB$, we have $p(E) = 1/4$. In Example 2 of slide number 24, we showed that $p(F) = 3/4$ and that $p(E \cap F) = 1/4$. But

$$p(E)p(F) = 1/4 \times 3/4 = 3/16$$

.

Therefore $p(E \cap F) \neq p(E)p(F)$, so the events E and F are not independent.

Discrete Probability Distributions (Recall MC3020)

In this section, we will be covering two significant probability distributions: the binomial and Poisson distributions. However, it is worth noting that you have already studied these distributions in the MC3020 course. As such, we will be revisiting these concepts here.

Random variable

A random variable is a quantitative variable whose values are determined by chance.

We use capital letter , like X, to denote the random variable and use small letter to list the possible values of the random variable.

Example

Suppose we toss a coin twice. Let X be the (random variable) number of heads. Sample space $S = \{HH, HT, TH, TT\}$ then we can construct a table

S	HH	HT	TH	TT
X	2	1	1	0

We also associate a probability with X attaining that value.

S	Prob	X
TT	1/4	0
TH	1/4	1
HT	1/4	1
HH	1/4	2

$\Rightarrow$

X	P(X=x)
0	1/4
1	1/2
2	1/4

Binomial Distribution

If a Bernoulli trial is repeated independently n times, we have a Binomial experiment. X is the number of successes in these n independent Bernoulli trials and p is the success probability in a trial. Then the distribution of X , with the probability mass function

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}; \quad x = 0, 1, 2, \cdots n$$

is known as the Binomial probability distribution. X is called a binomial random variable.

Poisson Distribution

X is the number of occurrences in a unit interval or space when the rate of occurrence is a fixed value $\lambda \geq 0$. Then the distribution of X, with the probability mass function

$$P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}; \quad x = 0, 1, 2, \cdots$$

is known as the Poisson probability distribution. X is called a Poisson random variable.

Please note that, the probability an event occurs in the interval is proportional to the length of the interval.

Examples

1. You are conducting a case-control study of smoking and lung cancer. If the probability of being a smoker among lung cancer cases is 0.6, what's the probability that in a group of 8 cases you have: (a) Less than 2 smokers? (b) More than 5?
2. In the manufacture of no deposit- no-return bottles, 5 percent of the bottles are defective. What is the probability in a sample of 7 bottles that there are, (a) no defectives (b) three defectives (c) more than three defectives.
3. You have submitted five proposals for upgrading the manufacturing facilities in your process area. From past experience you feel that the chance for any one project to be approved by the Finance Committee is 0.6. Accepting that $p = 0.6$ and that selection is a random event, what are the chances that at least one project will be approved?

Examples

1. Births in a hospital occur randomly at an average rate of 1.8 births per hour. What is the probability of observing 4 births in a given hour at the hospital? What is the probability of observing at least 2 births in a given hour at the hospital? What is the probability of observing 2 births in a given day at the hospital?
2. Airline passengers arrive randomly and independently at the passenger-screening facility at a major international airport. The mean arrival rate is 10 passengers per minute. What is the probability that three or fewer passengers will arrive in a 30 second period?
3. The transportation system at a new large airport is designed so that it will have one failure every ten days. What is the probability that it will not fail on the "Grand Opening Day"?

Bayes' Theorem

Suppose that E is an event from a sample space S and that $F_1, F_2, ..., F_n$ are mutually exclusive events such that $\cup_{i=1}^{n} F_i = S$. Assume that $p(E) \neq 0$ and $p(Fi) \neq 0$ for $i = 1, 2, \cdots n$. Then

$$P(F_j|E) = \frac{P(E|F_j)p(F_j)}{\sum_{i=1}^{n} P(E|F_i)p(F_i)}$$

Examples

1. Suppose that E and F are events in a sample space and $P(E) = 2/3, P(F) = 3/4$, and $P(F|E) = 5/8$. Find $P(E|F)$.

2. Three plants, C_1, C_2, and C_3, produce respectively, 10%, 50%, and 40% of a company's output. Although plant C_1 is a small plant, its manager believes in high quality and only 1% of its products are defective. The other two, C_2 and C_3, are worse and produce items that are 3% and 4% defective, respectively. All products are sent to a central warehouse. One item is selected at random and observed to be defective, then what is the conditional probability that it comes from plant C_1.(We define event D as defective)

3. Suppose that Roger selects a ball by first picking one of two boxes at random and then selecting a ball from this box. The first box contains three orange balls and four black balls, and the second box contains five orange balls and six black balls. What is the probability that Roger picked a ball from the second box if she has selected an orange ball?

Expected Value and Variance

The expected value of a random variable is the sum over all elements in a sample space of the product of the probability of the element and the value of the random variable at this element. Consequently, the expected value is a weighted average of the values of a random variable. The expected value of a random variable provides a central point for the distribution of values of this random variable.We can solve many problems using the notion of the expected value of a random variable, such as determining who has an advantage in gambling games and computing the average-case complexity of algorithms.

Another useful measure of a random variable is its variance, which tells us how spread out the values of this random variable are.We can use the variance of a random variable to help us estimate the probability that a random variable takes values far removed from its expected value.

Expected Values

The expected value, also called the expectation or mean, of the random variable X on the sample space S is equal to

$$E(X) = \sum_{\text{for all } x} xP(X = x)$$

The deviation of X at x is $x - E(X)$, the difference between the value of X and the mean of X.

Note that when the sample space S has n elements, then $S = x_1, x_2, \cdots x_n$,

$$E(X) = \sum_{i=1}^{n} x_i P(X = x_i)$$

.

Examples

1. Let X be the number that comes up when a fair die is rolled. What is the expected value of X?
2. A fair coin is flipped three times. Let S be the sample space of the eight possible outcomes, and let X be the random variable that assigns to an outcome the number of heads in this outcome. What is the expected value of X?
3. (Practice example) Proof that the expected value of the binomial random variable is np and the expected value of the binomial random variable is λ.

Linearity of Expectations

If $X_i, i = 1, 2, \cdots n$ with n a positive integer, are random variables on S, and if a and b are real numbers, then

1. $E(X_1 + X_2 + X_3 + \cdots X_n) = E(X_1) + E(X_2) + E(X_3) + \cdots E(X_n)$
2. $E(aX + b) = aE(X) + b$

Independent Random Variables

If X and Y are independent random variables on a sample space S, then

$$E(XY) = E(X)E(Y)$$

.

Remark:

We can use this result to check whether two random variables are independent without any proof in this course.

Variance

Let X be a random variable on a sample space S. The variance of X, denoted by $Var(X)$ or σ^2, is

$$Var(X) = \sum_{\text{for all } x} [x - E(X)]^2 \, P(X = x)$$

That is, $Var(X)$ is the weighted average of the square of the deviation of X, denoted by σ^2. The standard deviation of X, denoted σ, is defined to be $\sqrt{Var(X)}$.

Practice Problems

1. If X is a random variable on a sample space S, then

$$Var(X) = E(X^2) - (E(X))^2$$

.

2. If X is a random variable on a sample space S and $E(X) = \mu$, then

$$Var(X) = E((X - \mu)^2)$$

.

Practice Problems

1. What is the variance of the random variable X with $X = 1$ if a Bernoulli trial is a success and $X = 0$ if it is a failure, where p is the probability of success and q is the probability of failure?
2. What is the variance of the random variable X, where X is the number that comes up when a fair die is rolled?
3. If X and Y are two independent random variables on a sample space S, then $Var(X + Y) = Var(X) + Var(Y)$.
4. Show that if X and Y are independent random variables, then

$$Var(XY) = E(X)^2 Var(Y) + E(Y)^2 Var(X) + Var(X)Var(Y)$$

.

Don't hesitate to contact us if you have any questions about this course's teaching contents.

Also don't forget to check out the course page and Microsoft Team folder,

- course page https://mayooran1987.github.io/MC4010_E21/
- Microsoft Team folder link