1. Sample Space and Probability

Han-You Jeong

Networked Smart Systems Laboratory
School of Electrical and Computer Engineering
Pusan National University

Probability and Statistics

Outline

What is the probability?

- We use the concept of probability to discuss an uncertain situation.
- Probability shows either the frequency of occurrence or the subjective belief.
- The choices and actions of a rational person can reveal a lot about the inner-held subjective probabilities.
- Applications of probability include science, engineering, medicine, management, etc.
- The goal of this course is to develop the art of describing uncertainty in terms of probabilistic models, as well as the skill of probabilistic reasoning.
    - The models assign probabilities to collections (sets) of possible outcomes.

Set Theory

- A set is a collection of objects, which are the elements of the set.
  - If x is an element of set S, we write $x \in S$. Otherwise, we write $x \notin S$.
  - A set with no elements is called the empty set denoted by $\varnothing$.
  - If S contains a finite number of elements, say $x_1, x_2, \cdots, x_n$, we write it as
  $$S = \{x_1, x_2, \cdots, x_n\}.$$
  - Alternatively, all elements of set S have a certain property P, we write it as
  $$S = \{x | x \text{ satisfies } P\}.$$

- If all elements of set S can be enumerated in a list, we say that S is countable. Otherwise, we say that S is uncountable.
- If set S has a finite number of elements, we say that S is finite. Otherwise, we say that S is infinite.
  - countably finite (O), countably infinite (O), uncountably finite(X), uncountably infinite (O)
- If every element of a set S is also an element of a set T, we say that S is a subset of T, denoted by $S \subset T$.
- If $S \subset T$ and $T \subset S$, the two sets are equal.
- The universal set, denoted by Ω, contains all objects that could be conceivably be of interest in a particular context.

Set Operations

- The complement of a set S, denoted by S^c, is the set $\{x \in \Omega | x \notin S\}$ of all elements of Ω that do not belong to S.
- The union of two sets S and T, denoted by $S \cup T$, is the set of all elements that belong to S or T, i.e., $S \cup T = \{x | x \in S \text{ or } x \in T\}$.
- The intersection of two sets S and T, denoted by $S \cap T$, is the set of all elements that belong to both S and T, i.e., $S \cap T = \{x | x \in S \text{ and } x \in T\}$.
- Several sets are said to be disjoint if no two of them have a common element.
- A collection of sets is said to be a partition of a set S, if the sets in the collection are disjoint and their union is Ω.
- If x and y are two objects, we use (x, y) to denote the ordered pair of x and y.
  - one-dimension: $\Re$, two-dimension: $\Re^2$, three-dimension: $\Re^3$

Venn Diagrams

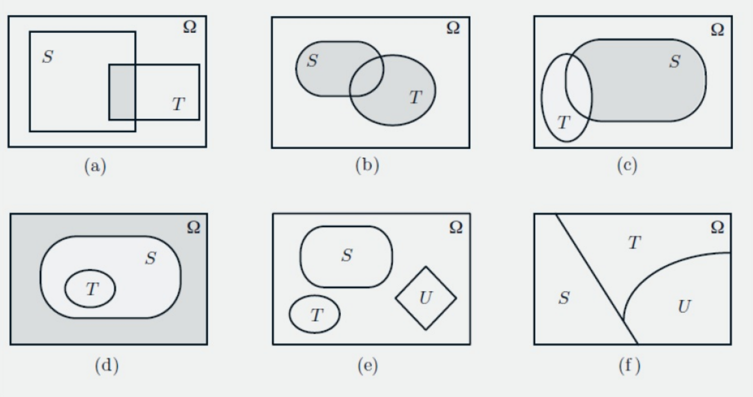


Figure 1.1: Examples of Venn diagrams. (a) The shaded region is $S \cap T$. (b) The shaded region is $S \cup T$. (c) The shaded region is $S \cap T^c$. (d) Here, $T \subset S$. The shaded region is the complement of S. (e) The sets S, T, and U are disjoint. (f) The sets S, T, and U form a partition of the set Ω.

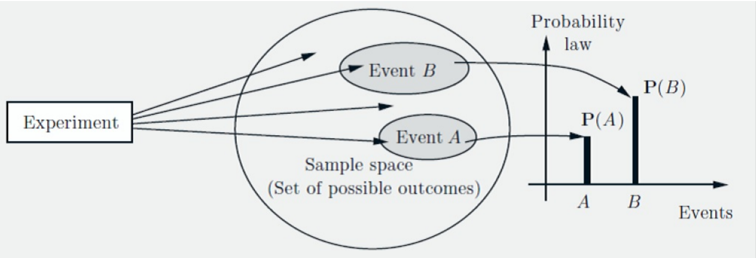
The Algebra of Sets

- Commutative Law: $S \cup T = T \cup S$
- Associative Law: $S \cup (T \cup U) = (S \cup T) \cup U$
- Distributive Law: $S \cup (T \cap U) = (S \cup T) \cap (S \cup U)$
- $(S^c)^c = S$, $S \cap S^c = \varnothing$, $S \cup \Omega = \Omega$, $S \cap \Omega = S$.
- De Morgan's Law

$$\left(\bigcup_n S_n\right)^c = \bigcap_n S_n^c, \quad \left(\bigcap_n S_n\right)^c = \bigcup_n S_n^c.$$

Elements of a Probability Model

- The sample space Ω is the set of all possible outcomes of an experiment.
- The probability law assigns to a set A of possible outcomes (called an event) a nonnegative number $P(A)$ (called probability of A) that encodes our knowledge or belief about the collected likelihood of the elements of A.



Choice of an Appropriate Sample Space

- Different elements of the sample space should be distinct and mutually exclusive, so that when the experiment is carried out there is a unique outcome.
- The sample space must be collective exhaustive in the sense that no matter what happens in the experiment, we always obtain an outcome that has been included in the sample space.

Ex 1.1 Consider ten successive coin tosses:

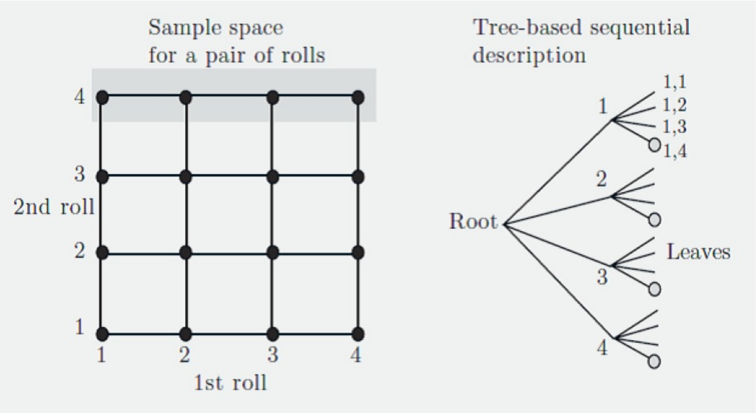
Game 1 We receive \$ 1 each time a head comes up.

Game 2 Starting with \$ 1, the dollar amount per toss is doubled each time a head consecutively comes up.

What is the sample space for each of the cases?

Sequential Model

- Many experiments have an inherently sequential character, e.g. rolling a four-sided dice twice, which can be represented by a tree-based sequential description.



Probability Axioms

- The probability law assigns to every event A, a number $P(A)$, called the probability of A, satisfying the following axioms.
  1. (Nonnegativity) $P(A) \geq 0$, for every event A.
  2. (Additivity) If A and B are two disjoint events, then the probability of their union satisfies

     $$P(A \cup B) = P(A) + P(B).$$

     More generally,

     $$P(A_1 \cup A_2 \cup \cdots) = P(A_1) + P(A_2) + \cdots.$$

  3. (Normalization) The probability of the sample space is equal to 1, that is $P(\Omega) = 1$.

Ex 1.2 Consider a single coin toss whose outcome is either head (H) or tail (T). Find the probability of the following events: $\{H, T\}$, $\{H\}$, $\{T\}$, $\varnothing$.

Discrete Models

- Discrete Probability Law
  - If the sample space consists of a finite number of possible outcomes, then the probability law is specified by the probabilities of the events that consist of a single element, i.e.,

$$P(\{s_1, s_2, \cdots, s_n\}) = P(s_1) + P(s_2) + \cdots + P(s_n).$$

- Discrete Uniform Probability Law
  - If the sample space consists of n possible outcomes which are equally likely, then the probability of any event A is given by

$$P(A) = \frac{\text{number of elements of } A}{n}.$$

Ex 1.3 Consider the experiment of rolling a pair of four-sided dice. What is the probability that

1. the sum of the rolls is even?
2. the first roll is equal to the second?
3. the first roll is larger than the second?
4. at least one roll is equal to four?

Continuous Models

- In probabilistic models with continuous sample spaces, the probabilities of the single-element events may not be sufficient to characterize the probability law.
  - We must consider an interval rather than an element in continuous sample spaces.

Ex 1.4 A wheel of fortune is continuously calibrated from 0 to 1, i.e., $\Omega = [0,\ 1]$. What is the probability that a wheel of fortune directs to a point in $[a,\ b]$, where $0 \leq a \leq b \leq 1$?

Ex 1.5 Romeo and Juliet have a date at a given time, each will arrive at the meeting place with a delay between 0 and 1 hour, with all pairs of delays being equally likely. The first to arrive will wait for 15 minutes and will leave if the other has not yet arrived. What is the probability that they will meet?

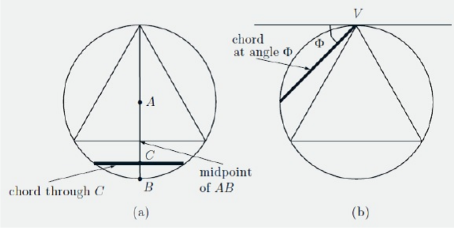
Properties of Probability Laws

- Consider a probability law, and let A, B, and C be events.
  - (a) If $A \subset B$, then $P(A) \leq P(B)$.
  - (b) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.
  - (c) $P(A \cup B) \leq P(A) + P(B)$.
  - (d) $P(A \cup B \cup C) = P(A) + P(A^c \cap B) + P(A^c \cap B^c \cap C)$.
- To generalize (c), we have

$$P(A_1 \cup A_2 \cup \cdots \cup A_n) \leq \sum_{i=1}^{n} P(A_i).$$

Models and Reality

- The framework of probability theory involves two distinct stages:
  1. We construct a probabilistic model by specifying a probability law on a suitably defined sample space.
  2. We derive the probabilities of certain events, or deduce some interesting probabilities.
- Bertrand's paradox
  - Consider a circle and an equilateral triangle inscribed in the circle. What is the probability that the length of randomly chosen chord of the circle is greater than the side of the triangle?



Conditional Probability

- Conditional probability provides us with a way to reason about the outcome of an experiment, based on partial information.
  - When you roll two dies, you are told that the sum of the two rolls is 9. How likely is it that the first roll was a 6?
  - How likely is it that a person has a certain disease given that a medical test was negative?
- Suppose that we know that the outcome is within some given event B. The conditional probability of A given B specifies the quantity of the likelihood that the outcome also belongs to some other given event A.

$$P(A|B) = \frac{\text{number of elements of } A \cap B}{\text{number of elements of } B} = \frac{P(A \cap B)}{P(B)}.$$

Here, we assume that $P(B) > 0$.

  - $P(\Omega|B) =?,\ P(A_1 \cup A_2|B) =?,\ P(A \cup C|B)\ ?\ P(A|B) + P(C|B)$.

Properties of Conditional Probability

- The conditional probability of an event A given an event B with $P(B) > 0$, is defined by

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

  where all properties of probability laws remain valid for conditional probability laws.

- Conditional probability can be viewed as a probability law on a new universe B, because all of the conditional probability is concentrated on B.

- If the possible outcomes are finitely many and equally likely, then

$$P(A|B) = \frac{\text{number of elements of } A \cap B}{\text{number of elements of } B}.$$

Some Examples

Ex 1.6 We toss a fair coin three times. We wish to find the conditional probability $P(A|B)$, when A and B are the events

$$A = \{\text{more heads than tails come up}\}, B = \{\text{1st toss is head}\}$$

Ex 1.7 A fair four-sided die is rolled twice. Let X and Y be the result of the 1st and the 2nd roll, respectively. We wish to determine the conditional probability $P(A|B)$, where

$$A = \{\max(X, Y) = m\}, B = \{\min(X, Y) = 2\},$$

and m takes each of the values 1, 2, 3, 4.

Using Conditional Probability for Modeling

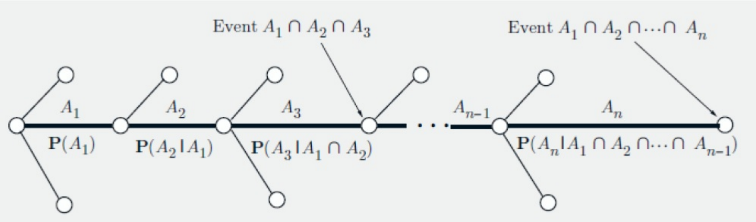
- When constructing probabilistic models for sequential experiments, it is convenient to first specify conditional probabilities and then use them to determine unconditional probabilities using $P(A \cap B) = P(B)P(A|B)$.

  Ex 1.9 If an aircraft is present in a certain area, a radar detects it and generates an alarm signal with probability 0.99. If the aircraft is not present, the radar generates a (false) alarm with probability 0.10. We assume that an aircraft is present with probability 0.05. What is the probability of no aircraft presence and a false alarm. What is the probability of aircraft presence and no detection?

Multiplication Rule

- Assuming that all of the conditioning events have positive probability, we have

$$P(\cap_{i=1}^{n} A_i) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \cdots P(A_n| \cap_{i=1}^{n-1} A_i).$$

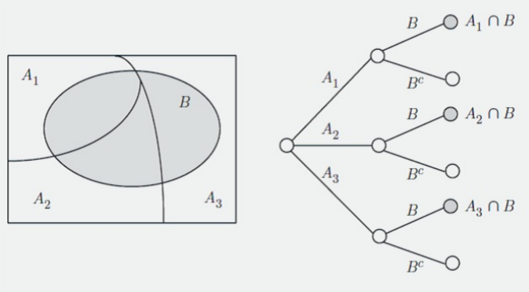


Ex 1.10 Three cards are drawn from an ordinary 52-card deck without replacement (drawn cards are not placed back in the deck). What is the probability that none of the three cards is heart?

Total Probability Theorem

- Let $A_1, \cdots, A_n$ be disjoint events that form a partition of the sample space (each possible outcome is included in exactly one of the events $A_1, \cdots, A_n$) and assume that $P(A_i) > 0$, for all i. Then, for any event B, we have

$$\begin{aligned} P(B) &= P(A_1 \cap B) + P(A_2 \cap B) + \cdots + P(A_n \cap B) \\ &= P(A_1)P(B|A_1) + \cdots + P(A_n)P(B|A_n). \end{aligned}$$



Some Examples

Ex 1.13 You enter a chess tournament where your probability of winning a game is 0.3 against half the players of type 1, 0.4 against a quarter of the players of type 2, and 0.5 against the remaining quarter of the players of type 3. You play a game against a randomly chosen opponent. What is the probability of winning?

Ex 1.14 You roll a fair four-sided die. If the result is 1 or 2, you roll once more but otherwise, you stop. What is the probability that the sum total of your rolls is at least 4?

Inference and Bayes' Rule

- Bayes' Rule
  - Let $A_1, A_2, \cdots, A_n$ be disjoint events that form a partition of the sample space, and assume that $P(A_i) > 0$, for all i. Then, for any event B such that $P(B) > 0$, we have

$$\begin{aligned} P(A_i|B) &= \frac{P(A_i)P(B|A_i)}{P(B)} \\ &= \frac{P(A_i)P(B|A_i)}{P(A_1)P(B|A_1) + \cdots + P(A_n)P(B|A_n)}. \end{aligned}$$

- Bayes' rule is often used for inference where the events $A_1, \cdots, A_n$ are associated with the causes and the event B represents the effect.
  - $P(A_i|B)$ is called the posterior probability of event A_i given the information, whereas $P(A_i)$ is called the prior probability.

Some Examples

Ex 1.16 Let us return to the radar detection problem of Ex 1.9, and wish to find the probability that aircraft is present given that the radar generates an alarm.

Ex 1.18 A test for a certain rare disease is assumed to be correct 95 % of the time: if a person has the disease, the test results are positive with probability 0.95, and if the person does not have the disease, the test results are negative with probability 0.95. A random person drawn from a certain population has probability 0.001 of having the disease. Given that the person just tested positive, what is the probability of having the disease?

Independence

- The conditional probability $P(A|B)$ is introduced to capture the partial information that event B provides about event A.
- We say that A is independent of B when the occurrence of B provides no such information and does not alter the probability that A has occurred, i.e.,

$$P(A|B) = P(A), \text{ and } P(A \cap B) = P(B)P(A|B) = P(A)P(B).$$

  - Two disjoint events A and B with $P(A) > 0$ and $P(B) > 0$ are never independent. Why?

Ex 1.19 Consider an experiment of rolling two four-sided die in which all 16 outcomes are equally likely happen.

1. Are the events $A_i = \{\text{1st roll results in } i\}$, $B_j = \{\text{2nd roll results in } j\}$ independent?
2. Are the events $A = \{\text{maximum of the two rolls is 2}\}$, $B = \{\text{minimum of the two rolls is 2}\}$ independent?

Conditional Independence

- Given an event C, the events A and B are called conditionally independent if

$$\begin{aligned} P(A \cap B|C) &= \frac{P(A \cap B \cap C)}{P(C)} \\ &= \frac{P(C)P(B|C)P(A|B \cap C)}{P(C)} \\ &= P(B|C)P(A|B \cap C) \\ &= P(B|C)P(A|C). \end{aligned}$$

- Conditional independence is the same as the condition $P(A|B \cap C) = P(A|C)$.
- Notice that independence of two events A and B with respect to the unconditional probability law, does not imply conditional independence, and vice versa.

An Example

Ex 1.20 Consider two independent fair coin tosses with events

$$H_1 = \{\text{1st toss is a head}\},$$

$$H_2 = \{\text{2nd toss is a head}\},$$

$$D = \{\text{the two tosses have different results}\}.$$

Are the events H_1 and H_2 independent? How about the conditional events $H_1|D$ and $H_2|D$?

Independence of Collection of Events

- We say that the events $A_1, A_2, \cdots, A_n$ are independent if

$$P\left(\bigcap_{i\in S} A_i\right) = \prod_{i\in S} P(A_i), \text{ for every subset } S \text{ of } \{1, 2, \cdots, n\}.$$

- If any two events are independent, we say that the events are pairwise independent.

Ex 1.22 In the Ex 1.20, are the three events H_1, H_2, and D (pairwise) independent?

Ex 1.23 Consider two independent rolls of fair six-sided die, and the following events:

$$A = \{\text{1st roll is 1, 2, or 3}\}$$

$$B = \{\text{1st roll is 3, 4, or 5}\}$$

$$C = \{\text{the sum of the two rolls is 9}\}$$

Then, are these three events (pairwise) independent?

Understanding of the Independence

- Independence means that the occurrence or non-occurrence of any number of the events from that collection carries no information on the remaining events or their complements.
  - For example, if the events A_1, A_2, A_3, A_4 are independent, one obtains the following relations:

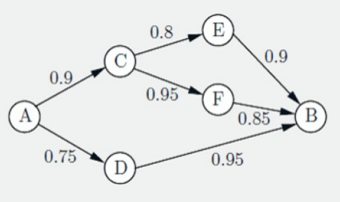
$$P(A_1 \cup A_2 | A_3 \cap A_4) = P(A_1 \cup A_2),$$

$$P(A_1 \cup A_2^c | A_3^c \cap A_4) = P(A_1 \cup A_2^c).$$

Reliability

- In probabilistic models of complex systems involving several components, it is convenient to assume that the behaviors of the components are uncoupled (independent).

Ex 1.24 A computer network connects two nodes A and B through intermediate nodes C, D, E, F, as shown below. In addition, there is a given probability p_{ij} that the link from i to j is up. We assume that link failures are independent of each other. What is the probability that there is a path connecting A and B in which all links are up?



Independent Trials and the Binomial Probabilities

- We say that we have a sequence of independent trials, if an experiment involves a sequence of independent but identical stages.
- We say that we have a sequence of independent Bernoulli trials in the special case where there are only two possible results at each stage.
  - At each stage, the probability that an event happens is equal to p.
  - Let us define the probability $p(k) = P(k \text{ heads come up in an } n\text{-toss sequence})$, then probability $p(k)$ is given by
  $$p(k) = \binom{n}{k} p^k (1-p)^{n-k},$$
  where $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ are known as the binomial coefficients, while the probabilities $p(k)$ are known as the binomial probabilities.
- Notice that the binomial probabilities $p(k)$ must add to 1, thus showing the binomial formula,
$$\sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} = 1.$$

An Example

Ex 1.25 An internet service provider has installed c modems to serve n dial-up customers. It is estimated that at a given time, each customer will need a connection with probability p, independently of the others. What is the probability that there are more customers needing a connection than there are modems?

Counting

- The calculation of probabilities often involves counting the number of outcomes in various events.
  - When the sample space Ω has a finite number of equally likely outcomes,
  $$P(A) = \frac{\text{number of elements of } A}{\text{number of elements of } \Omega}.$$
  - When we want to calculate the probability of an event A with a finite number of equally likely outcomes, each of which has an already known probability p,
  $$P(A) = p \cdot (\text{number of elements of } A)$$
- The art of counting constitutes a large portion of the field of combinatorics.

The Counting Principle

- The counting principle is based on a divide-and-conquer approach, whereby the counting is broken down into stages through the use of tree.
- Consider a process that consists of r stages. Suppose that
  1. There are n_1 possible results at the first stage.
  2. For every possible result at the first stage, there are n_2 possible results at the second stage.
  3. For any sequence of possible results at the first $i-1$ stages, there are n_i possible results at the ith stage. Then, the total number of possible results of the r-stage process is $n_1 n_2 \cdots n_r$.

Ex 1.26 A local telephone number is a 7-digit sequence, but the first digit must be different from 0 or 1. How many distinct telephone numbers are there?

Ex 1.27 Consider an n-element set. How many subsets does it have?

Permutations

- In *k*-permutations, we wish to count the number of different ways that we can pick k out of n objects, and arrange them in a sequence, which is given by

$$n(n-1)\cdots(n-k+1) = \frac{n(n-1)\cdots(n-k+1)(n-k)\cdots 2\cdot 1}{(n-k)\cdots 2\cdot 1} = \frac{n!}{(n-k)!}.$$

Ex 1.28 Count the number of alphabetic words that consists of four distinct letters.

Ex 1.29 You have n_1 classic CDs, n_2 rock music CDs, and n_3 country music CDs. In how many different ways can you arrange them so that the CDs of the same type are contiguous?

Combinations

- In combinations, we wish to count the number of k-element subsets of a given n-element set, where there is no ordering of the selected elements.
  - Each combination is associated with $k!$ duplicate k-permutations, thus we have
$$\binom{n}{k} = \frac{n!}{(n-k)!k!}.$$
  - From the binomial formula with $p = 0.5$, we have
$$\sum_{k=0}^{n} \binom{n}{k} = 2^n.$$

Ex 1.30 Count the number of combinations of two out of four letters A, B, C, and D.

Partitions

- A combination can be viewed as a partition of the set in two: one part contains k elements and the other contains the remaining $n-k$.
- We consider partitions of the set into r disjoint subsets, with the ith subset containing exactly n_i elements.
  - The total number of choices is

$$\begin{aligned}
&\binom{n}{n_1}\binom{n-n_1}{n_2}\binom{n-n_1-n_2}{n_3}\cdots\binom{n-n_1-\cdots-n_{r-1}}{n_r} \\
&= \frac{n!}{n_1!(n-n_1)!}\cdot\frac{(n-n_1)!}{n_2!(n-n_1-n_2)!}\cdots\frac{(n-n_1-\cdots-n_{r-1})!}{n_r!(n-n_1-\cdots-n_{r-1}-n_r)!} \\
&= \frac{n!}{n_1!n_2!\cdots n_r!} = \binom{n}{n_1, n_2, \cdots, n_r}.
\end{aligned}$$

Ex 1.32 How many different words (letter sequences) can be obtained by rearranging the letters in the word TATTOO?

Homework # 1

- Section 1.1: Problem 1
- Section 1.2: Example 1.5, Problem 8
- Section 1.3: Example 1.9, Problem 17
- Section 1.4: Example 1.18, Problem 24
- Section 1.5: Example 1.24, Problem 36
- Section 1.6: Example 1.29, Problem 54

Thank You