Stat 115: Probability

Prof. Josefina V. Almeda

Development of Probability Theory

- French mathematicians Pierre de Fermat (1601-1665) and Blaise Pascal (1623-1662) started formulating the fundamental principles of probability theory.

- Came about as answers to questions of gamblers on games involving dice or cards

10.1 Probability Models

Definition 10.1: Abstract Model

An abstract model is a description of the essential properties of a phenomenon that is formulated in mathematical terms.

Example: area of a rectangle

10.1 Probabilistic Models

Definition 10.2: Deterministic Model

A deterministic model is a type of abstract model that describes a phenomenon through known relationships among the states and events, in which a given input will always produce the same output.

Notes for Deterministic Model

- Given the same input, the model will always produce the same output.
- This model does not leave any room for random variation.
- We cannot describe the outcomes in a game of chance using this type of model.

10.1 Probabilistic Models

Definition 10.3: Probabilistic/Stochastic Model

It is a type of abstract model that describes a phenomenon by assigning a likelihood of occurrence to the different possible outcomes of the process.

Notes on Probabilistic Model

- Example: toss of a coin
- No matter how many times we repeat the process, it is impossible to predict with certainty what the next outcomes will be.
- Given the same inputs, the output may be different. (example: experiments)
- This is the type of model that we use in Inferential Statistics. It helps us predict with some degree of confidence the future outcomes.

Basic Concepts of Probability

Definition 10.4: Random Experiment

It is a process that can be repeated under similar conditions but whose outcome cannot be predicted with certainty beforehand.

Examples: toss of a coin, toss of a die

Note on Random Experiments

- In inferential statistics, the process of selecting a sample of size n from a population size N using probability sampling is one of the random experiments of interest.
- It is just like selecting n cards at random from a deck of N = 52 cards.
- Even if we use exactly the same sample selection procedure, there is no way we can predict, without any error, what the composition of the next sample will be.

Features of a Random Experiment

- All outcomes are known in advance.
- The outcome of any one trial cannot be predicted with certainty.
- Trials can be repeated under identical conditions.

Basic Concepts of Probability

Definition 10.5: Sample Space

The sample space, denoted by Ω (Greek letter, omega), is the collection of all possible outcomes of a random experiment. An element is called a sample point.

Two Ways of Specifying a Set

1. Roster Method

   - listing down all the elements belonging in the set then enclosing them in braces.

2. Rule Method

   - stating a rule that the elements must satisfy in order to belong in the set then enclosing this rule in braces.

Example of Roster and Rule Method

Example 1: Toss a coin once.

Roster Method: $\Omega = \{H, T\}$

Rule Method : $\Omega = \{x/x \in \{H,T\}\}$

Example 2: Toss a coin twice.

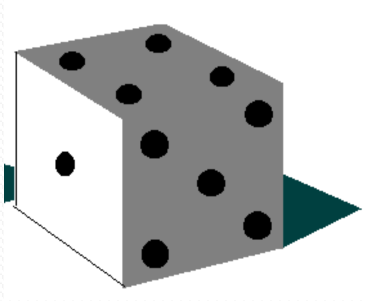
Roster Method: $\Omega = \{HH, HT, TH, TT\}$

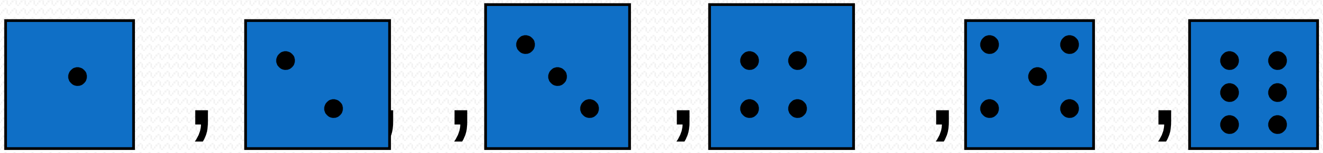
Rule Method : $\Omega = \{(x,y)/\ x \in \{H,T\} \text{ and } y \in \{H,T\}\}$

NOTE: Tree diagram

ILLUSTRATION

Rolling a die and observing the number of dots on the upturned face



Ω={  }

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

Basic Concepts of Probability

Definition 10.6: Event

An event is a subset of the sample space whose probability is defined. We say that an event occurred if the outcome of the experiment is one of the sample points belonging in the event; otherwise, the event did not occur.

We use any capital Latin letter to denote an event.

ILLUSTRATION

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

An event of observing odd-number of dots in a roll of a die

$E_1 = \{ 1, 3, 5\}$

An event of observing even-number of dots in a roll of a die

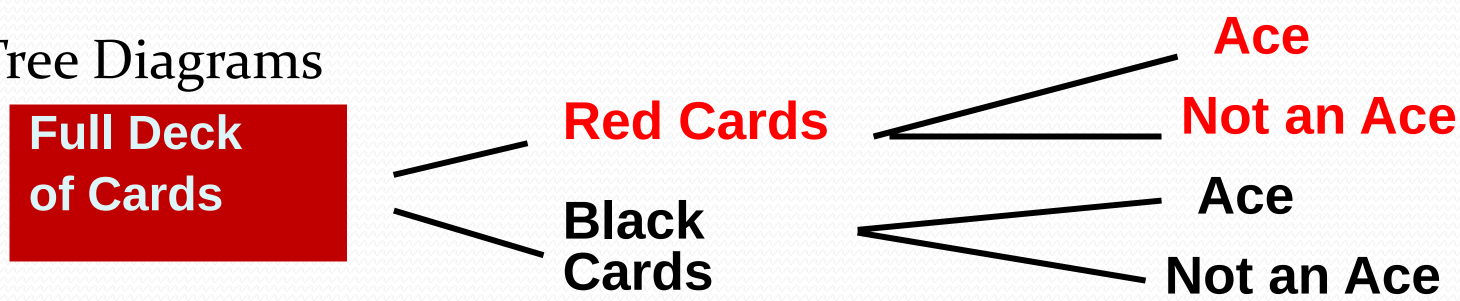
$E_2 = \{ 2, 4, 6\}$

Visualizing Events

- Contingency Tables

	Ace	Not Ace	Total
Black	2	24	26
Red	2	24	26
Total	4	48	52

- Tree Diagrams



Basic Concepts in Probability

Definition 10.7: Impossible Event and Sure Event

The impossible event is the empty set, Ø. The sure event is the sample space, Ω.

Reading Assignment: p. 289 to 291

Examples of Events using Venn Diagram

- Universal set
- The set A
- A complement
- A U B
- A U B U C
- A $\cap$ B
- A $\cap$ B$\cap$ C
- A $\cap$ B^c
- A and B are mutually exclusive events

Basic Concepts of Probability

Definition 10.8: Mutually Exclusive Events

Two events A and B are mutually exclusive events if and only if $A \cap B = \emptyset$; that is, A and B have no elements in common.

Mutually Exclusive Events

Two events are ***mutually exclusive*** if the two events cannot occur simultaneously.

Example:

Coin toss: either a head or a tail, but not both. The events head and tail are mutually exclusive.

Examples of Mutually Exclusive Events

- Events A and A^c
- Events A and $B \cap A^c$
- Events $A \cap B$ and $A \cap B^c$
- Any event A and Ø

Basic Concepts of Probability

Definition 10.9: Probability of an event A (by Andrey Kolmogorov (1903-1987)

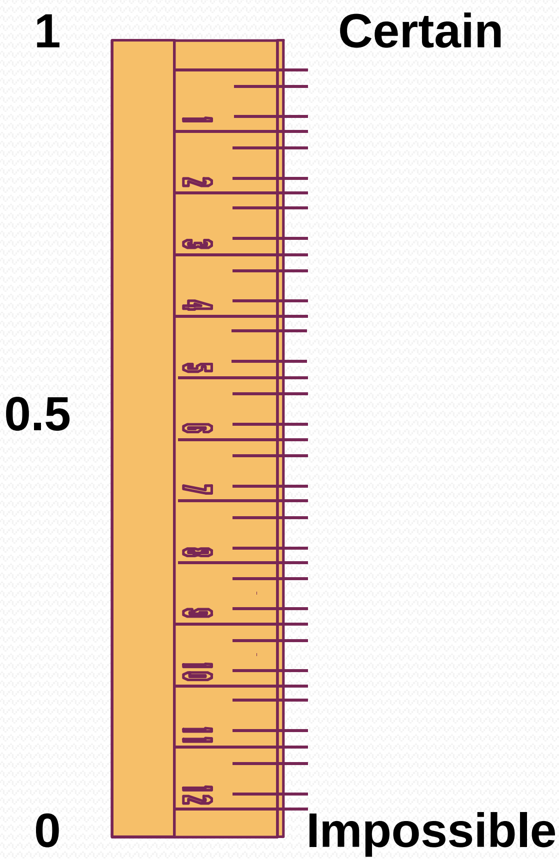
The probability of an event A, denoted by P(A), is a function that assigns a measure of chance that event A will occur and must satisfy the following properties:

a) $0 \leq P(A) \leq 1$ for any event A
b) $P(\Omega) = 1$
c) Finite Additivity. If A can be expressed as the union of n mutually exclusive events, that is, $A = A_1 \cup A_2 \cup \ldots \cup A_n$, then $P(A_1) + P(A_2) + \ldots + P(A_n)$.

PROBABILITY

- The numerical measure of the likelihood that an event will occur
- Between 0 and 1

Note: Sum of the probabilities of all mutually exclusive and collective exhaustive events is 1



Exercise

Find the errors in each of the following assignments of probabilities.

a. The probabilities that a student will have 0, 1, 2, 3, or 4 or more mistakes are 0.41, 0.37, 0.24, 0.13, and -0.15 respectively.

b. The probability that it will rain tomorrow is 0.46 and the probability that it will not rain tomorrow is 0.55.

c. The probabilities that an automobile salesperson will sell 0, 1, 2, or 3 items on any given day are, 0.24, 0.43, 0.26, and 0.21.

d. On a single draw from a deck of playing cards the probability of selecting a heart is 1/4, the probability of selecting a black card is ½, and the probability of selecting both a heart and a black card is 1/8.

Assignment

- Do pp. 295-296 (Exercises for section 10.2)

10.3 Approaches to Assigning Probabilities

- Subjective

  confident student views chances of passing a course to be near 100 %
- Logical

  symmetry/equally likely: coin, dice, cards etc. (A PRIORI assignment)
- Empirical

  chances of rain 75 % since it rained 15 out of past 20 days (A POSTERIORI)

Approaches to Assigning Probabilities

Definition 10.10: A Priori or Classical Probability

(Gerolamo, Cardano (1501-1576)

The method of using a priori or classical probability assigns probabilities to events before the experiment is performed using the following rule:

If an experiment can result in any one of N different equally likely outcomes, and if exactly n of these outcomes belong to event A, then

$$P(A) = \frac{no.of\ elements\ in\ A}{no.of\ elements\ in\ \Omega} = \frac{n}{N}$$

Notes on A Priori

- Its use is restricted to experiments whose sample space contains equiprobable outcomes, and consequently, the sample space must have only a finite number of sample points.

Steps in Assigning Probabilities Using A Priori

Step 1: Specify the sample space. Make sure that the outcomes are equiprobable

Step 2: Specify event A whose probability you are interested in.

Step 3: Count the number of samples points in Ω and denote this by N. Count the number of sample points in event A and denote this by n.

Step 4: Compute for the probability of event A using the formula, $P(A) = n/N$.

Examples of A Priori:

1. A fair coin is tossed 3 times. Find the probability of the following events:

   a) A = event of observing tails in the first 2 tosses

   b) B = event of observing exactly 2 tails

   c) C = event of observing at most 1 tail

2. In a deck of cards, find the probability of the following events:

   a) A = event of getting a King

   b) B = event of getting a spade

   c) C = event of getting a card lower than 5

Note on Classical Definition of Probability:

- This allows us to view proportions in terms of probabilities.
- We define the sample space as the collection of elements in the population.
- Let A = event that the selected element possesses the characteristic of interest.
- Thus, by classical definition of probability:

$$P(A) = \frac{no.of\ elements\ in\ A}{no.of\ elements\ in\ the\ population}$$

= proportion of elements possessing the characteristic of interest

Example of Proportions in terms of Probabilities

- Let A = event of selecting a student who uses the
  Facebook more than 3 hours a day

- Let B = event of selecting a student who uses the
  Globe line

Approaches to Assigning Probabilities

Definition 10.11: Posteriori or Relative Frequency

The method of using a posteriori or relative frequency assigns probabilities to events by repeating the experiment a large number of times and using the following rule:

If a random experiment is repeated many times under uniform conditions, use the **empirical probability** of event A to assign its probability as follows:

$$\text{empirical } P(A) = \frac{\textit{no. of times event A occurred}}{\textit{no. of times experiment was repeated}}$$

Notes on A Posteriori

- The A Posteriori of the probability of event A is the **limiting value of its empirical probability or relative frequency of occurrence of event A if we repeat the process endlessly and under uniform conditions.**

- A Priori probabilities uses deduction by assuming equal case occurrence

- A Posteriori probabilities uses induction from relative frequencies.

Notes on A Posteriori

- Advantage of Using A POSTERIORI: not restricted to random experiments that generate a sample space containing equiprobable outcomes.

- This approach can now be used to assign probabilities to events that arise from die-throwing experiments where the die is NOT fair; or from coin-tossing experiments where the coin is NOT balanced.

Approaches to Assigning Probabilities

Definition 10.12: Subjective probability

Subjective probability assigns probabilities to events by using intuition, personal beliefs, and other indirect information.

HOMEWORK: Do exercises 1 and 3 for section 10.3, p. 302

RULES OF COUNTING

Theorem 10.1: Basic Principle of Counting

Suppose an experiment can be performed in two stages. If there are n distinct possible outcomes in the first stage of the experiment and if, for each outcome of the first stage, there are m distinct possible outcomes in the second stage, then there are **n x m** possible outcomes of this experiment.

Rules of Counting

Theorem 10.2: Generalized Basic Principle of Counting (Multiplication Rule)

Suppose an experiment can be performed in k stages. If there are n_1 distinct possible outcomes in the first stage, and if for each of these n_1 outcomes there are n_2 distinct possible outcomes in the second stage, and if for each of the $\boldsymbol{n_1 x n_2}$ outcomes of the first 2 stages, there are n_3 distinct possible outcomes in the third stage; and continuing in this manner, until we reach the last stage where there are n_k distinct possible outcomes for each of the outcomes of the first (k-1) stages, then there are $\boldsymbol{n_1 \, x \, n_2 \, x \, n_3 x \ldots x \, n_k}$ possible outcomes of the experiment

Examples Using Generalized Basic Principle of Counting

1. How many sample points are there in the sample space when a coin is tossed once, twice, and thrice.
2. How many sample points are there in the sample space when a die is tossed once and twice?
3. How many even 3-digit numbers can be formed from the digits 1,2,5,6, and 9 if each digit can be used only once?
4. How many ways can a 10-question true-false examination be answered?

Rules of Counting

Definition 10.13: r-permutation

An **r-permutation** of set Z is an ordered arrangement of *r* distinct elements selected from the set Z. It can be represented by an ordered *r*-tuple with distinct coordinates. If set Z contains *n* distinct elements then the number of r-permutations of set Z is denoted by P(n,r) or $_{n}P_{r.}$, read as “permutation n taken r”.

Rules of Counting

Definition 10.14: r-combination

An **r-combination** of set Z is a subset of set Z that contains *r* distinct elements. If set Z contains *n* distinct elements then the number of r-combinations of set Z is denoted by C(n,r) or , read as “n taken r”.

Examples of r-permutation and r-combination

1. Suppose Z = { A, B, C }. List down all the possible 3-permutations of Z. List down all the possible 3-combinations of Z.

2. Suppose Z = { A, B, C, D }. List down all the possible 4-permutations of Z. List down all the possible 4-combinations of Z.

Rules of Counting

Definition 10.15: factorial notation

The **factorial notation** is a compact representation for the product of the first n consecutive positive integers. It is denoted by n! (read as “n factorial”) and $n!=(n)x(n-1)x(n-2)x\ldots x(2)x(1)$ where n is a positive integer. We also define $0!=1$.

Rules of Counting

Theorem 10.3: The number of distinct r-permutations that we can form from the n distinct elements of the set Z is

$$P(n,r) = (n)x(n-1)x(n-2)x\ldots x(n-r+1) = \frac{n!}{(n-r)!}$$

As a corollary, the number of permutations of all the *n* distinct elements of set Z is n!. We can derive this by taking r=n so that P(n,n)= n! .

Examples of Permutation:

1. Two lottery tickets are drawn from 20 for the first and second prize. Find the number of sample points in the sample space.

2. In how many ways can the 5 starting positions on a basketball team be filled with 8 men who can play any position?

Examples of Permutation:

3. A classroom has 6 rows of desks with 8 desks in a row. A class consists of 10 boys and 15 girls.

   a) How many ways can the teacher select and arrange 8 students in this class who will occupy the first row?

   b) Suppose the teacher does not want any boy to occupy the first row. How many ways can the teacher select and arrange 8 students who will occupy the first row?

Solution To Example no. 3:

Solution: a) There are 8 stages in the experiment. The i^{th} stage is the selection of the student who will occupy the i^{th} position in the first row, i=1,2,…,8. Since a student cannot occupy two positions at the same time, whoever was selected in the previous stage cannot anymore be selected in the current stage. We can use Theorem 10.3 to answer this question where r=8 and n=25 because the teacher is choosing 8 distinct students from 25 students. Thus, the answer is P(25,8)= 43,609,104,000.

Solution To Example no. 3:

Solution:

b) Using Theorem 10.3 again where r=8 but this time n=15 (since the teacher is choosing from 15 girls only) gives us the answer,

$$P(15,8)= 259{,}459{,}200$$

Rules of Counting

Theorem 10.4: The number of distinct *r*-combinations that can be formed from the *n* distinct elements of set Z is:

$$C(n,r)= P(n,r)/r! = n!/(n-r)!r!$$

Examples of Combinations

1. In a statistics exam, a student has a choice of 8 questions out of 10. In how many ways can he choose a set of 8 questions if he chooses arbitrarily?
2. Find the number of ways of selecting the 6 winning numbers in the original version of the game of lotto.
3. Consider the game of poker where a player is given 5 cards.

a) How many 5-card poker hands (unordered) are there?

b) How many of these 5-card poker hands contain exactly 3 hearts?

Solution to Example No. 3:

a) In this problem, we ignore the order in which the 5 cards were dealt. The answer to the question is C(52,5). By Theorem 10.4, there are as many as poker hands 2,598,960 poker hands.

b) The experiment can be divided into 2 stages: (i) the selection of the hearts; and then, (ii) the selection of the non-hearts. Then by the basic principle of counting, there are n_1 x n_2 poker hands containing exactly 3 hearts where n_1 = number of ways of selecting 3 hearts and n_2 = number of ways of selecting 2 non-hearts. By Theorem 10.4, n1 = C(13,3) = 13!/10!3! = 286 and n2 = C(39,2) = 39!/37!2! = 741. and . Thus, the answer is (286)(741)=211,926.

Theorem : if n objects are not anymore distinct from each other.

- Theorem 10.5: The number of distinct ways of arranging n objects of which n_1 are of one kind, n_2 are of a second kind, …, n_k are of a k^{th} kind is:

$$\frac{n!}{n_1!\,x\,n_2!\,x \ldots x\,n_k!} \quad where \ \sum_{i=1}^{k} n_i = n$$

Example for Theorem 10.5:

Consider our favorite word, “STATISTICS”. How many distinct ways can we arrange the letters contained in this word.

Solution:

Example for Theorem 10.5

In how many different ways can 3 red, 4 yellow, and 2 blue bulbs be arranged in a string of Christmas tree lights with 9 sockets?

Theorem 10.6: General Partitioning

The number of distinct ways of grouping n distinct objects into k groups such that n_1 objects belong in the first group, n_2 objects belong in the second group,…, n_k objects belong in the k^{th} group is:

$$\frac{n!}{n_1! \, x \, n_2! \, x \ldots x \, n_k!} \quad where \ \sum_{i=1}^{k} n_i = n$$

- **Example 10.21:** How many ways can we assign twenty new applicants into the 5 committees of an organization so that each committee will get 4 new applicants each?
- *Solution:*

- Do Exercises for Section 10.4

10.5 Properties of a Probability Function

Theorem 10.7:

If A is an event then $P(A^c) = 1 - P(A)$.

Theorem 10.8:

If A and B are events then $P(A \cap B^c) = P(A) - P(A \cap B)$.

Theorem 10.9. Additive Law of Probability

If A and B are events then $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Example for Properties of a Probability Function

- **Example 10.22:** Suppose A and B are events for which it is known that $P(A)=0.6$, $P(B)=0.7$ and $P(A\cap B)=0.4$. We can then compute the probabilities of other events using Theorems 10.7 – 10.9.
- $P(A\cup B) = P(A) + P(B) - P(A\cap B) = 0.6 + 0.7 - 0.4 = 0.9$.
- $P(A\cap B^c) = P(A) - P(A\cap B) = 0.6 - 0.4 = 0.2$.
- $P(B\cap A^c) = P(B) - P(A\cap B) = 0.7 - 0.4 = 0.3$
- $P((A\cap B)^c) = 1 - P(A\cap B) = 1 - 0.4 = 0.6$
- $P((A\cup B)^c) = 1 - P(A\cup B) = 1 - 0.9 = 0.1$

Remarks:

- The method of computing for the probabilities of events using these theorems is called the *event-composition method.*

- As the name implies, the probabilities are computed by expressing the event of interest as a composition of other events.

- **Example 10.23:** The probability that a randomly selected student passes Stat 101 is 0.60, and the probability that he passes English 11 is 0.85. If the probability that he passes at least one of the two courses is 0.95,
- what is the probability that the selected student passes both courses?
- what is the probability that he will fail both Stat 101 and English 11?

Solution to Example:

Let A=event that selected student passes Stat 101

B=event that selected student passes English 11

Given: $P(A)=0.60$ $P(B)=0.85$ $P(A \cup B)=0.95$

Find $P(A \cap B)$. Using Theorem 10.9, we have

$P(A \cap B) = P(A) + P(B) - P(A \cup B) = 0.60 + 0.85 - 0.95 = 0.5$

Find $P(A^c \cap B^c) = P(A \cup B)^c$. Using Theorem 10.7, we have

$P(A \cup B)^c = 1 - P(A \cup B) = 1 - 0.95 = 0.05$

Example:

In the toss of a fair coin 4 times, what is the probability of no head in the toss? At least one head?

- HW: Do exercises for Section 10.5

10.6 Conditional Probability

- Definition 10.16
- Let A and B be two events where P(B)>0. The **conditional probability** of event A given the occurrence of event B, denoted by P(A|B) (read as "probability of A given B) is:

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}.$$

- **Example :** The probability that a regularly scheduled flight departs on time is $P(D) = 0.83$, the probability that it arrives on time is $P(A) = 0.92$, and the probability that it departs and arrives on time is $P(D \cap A) = 0.78$. Find the probability that a plane
- (a) arrives on time given that it departed on time, and
- (b) departed on time given that it has arrived on time.

Special Theorems on Conditional Probabilities

- The conditional probability is also a probability function so it enjoys all of the properties of a probability function. Specifically, suppose B is an event satisfying the condition that P(B)>0 then,
- $P(\varnothing|B) = 0$.
- If $A_1, A_2, \ldots,$ and A_n are mutually exclusive events then

$$P(A_1 \cup A_2 \cup \ldots \cup A_n|B)=P(A_1|B) + P(A_2|B) + \ldots + P(A_n|B).$$

Special Theorems on Conditional Probabilities

- The conditional probability is also a probability function so it enjoys all of the properties of a probability function. Specifically, suppose B is an event satisfying the condition that $P(B)>0$ then,
- If A is an event then $P(A^c|B) = 1 - P(A|B)$.
- If A_1 and A_2 are events then $P(A_1 \cap A_2^c|B) = P(A_1|B) - P(A_1 \cap A_2|B)$.
- If A_1 and A_2 are events then $P(A_1 \cup A_2|B) = P(A_1|B) + P(A_2|B) - P(A_1 \cap A_2|B)$.

Example:

- **Example 10.29:** In example 10.23, the probability that the randomly selected student passes Stat 101 is 0.60, the probability that he passes English 11 is 0.85, and the probability that he passes at least one of the two courses is 0.95. What is the probability that the selected student fails English 11 given that he failed Stat 101?

Solution to Example 10.29

Solution:

let A =event that selected student passes Stat 101

B =event that selected student passes English 11

We want $P(B^c|A^c)$.

In example 10.25, we have already computed the probability that the selected student passes English 11 given that he failed in Stat 101 as $P(B|A^c)= 0.875$.

Then, we can just use the property of the conditional probability as a probability function to compute for $P(B^c|A^c)$.

$P(B^c|A^c) = 1 - P(B|A^c) = 1 – 0.875 = 0.125$.

Theorem 10.10. Theorem of Total Probabilities

- If $\{B_1, B_2, \ldots, B_n\}$ is a collection of mutually exclusive events wherein each event has a non-zero probability and

  $\Omega = B_1 \cup B_2 \cup \ldots \cup B_n$, then for any event A,

  $$P(A) = \sum_{j=1}^{n} P(A \mid B_j) P(B_j)$$

that is,

$$P(A) = P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + \ldots + .P(A|B_n)P(B_n)$$

Corollary to Total Probabilities

Since events B and B^c are mutually exclusive events and $B \cup B^c = \Omega$ then applying the Theorem of Total Probabilities, we have:

$$P(A) = P(A|B)P(B) + P(A|B^c)P(B^c).$$

Example 1: Theorem of Total Probability

Three members of a private country club have been nominated for the office of president. The probability that Mr. Acusta will be elected is 0.3, the probability that Mr. Bautista will be elected is 0.5, and the probability that Ms. Catral will be elected is 0.2. Should Mr. Acusta be elected, the probability for an increase in membership fees is 0.8. Should Mr. Bautista or Ms. Catral be elected, the corresponding probabilities for an increase in fees are 0.1 and 0.4. What is the probability that there will be an increase in membership fees?

Solution: Consider the following events:

A: membership fees are increased,

B_1: Mr. Acusta is elected

B_2: Mr. Bautista is elected

B_3: Ms. Catral is elected

Applying the rule of elimination, we can write

$$P(A) = P(B_1)P(A/B_1) + P(B_2)P(A/B_2) + P(B_3)P(A/B_3)$$

Solution:

We find that the three branches given the probabilities

$P(B_1)P(A/B_1) = (0.3)(0.8) = 0.24,$

$P(B_2)P(A/B_2) = (0.5)(0.1) = 0.05$

$P(B_3)P(A/B_3) = (0.2)(0.4) = 0.08$

And hence

$P(A) = 0.24 + 0.05 + 0.08 = 0.37$

Continuation of Problem:

- Suppose that we now consider the problem of finding the conditional probability P(B3/A). In other words, if it is known that membership fees have increased, what is the probability that Ms. Catral was elected president of the club?

- Questions of this type can be answered by using the following theorem called Bayes' rule:

Theorem 10.11: Bayes' Theorem

- If $\{B_1, B_2, \ldots, B_n\}$ is a collection of mutually exclusive events wherein each event has a non-zero probability and $\Omega = B_1 \cup B_2 \cup \ldots \cup B_n$, then for any event A for which $P(A)>0$,

$$P(B_k \mid A) = \frac{P(A \mid B_k)P(B_k)}{\sum_{j=1}^{n} P(A \mid B_j)P(B_j)}$$

Corollary to Bayes' Theorem

- Since events B and B^c are mutually exclusive events wherein each event has a nonzero probability and
  B U B^c = Ω then applying Bayes' Theorem, we have

$$P(B|A) = \frac{P(A|B)P(B)}{P(A|B)P(B) + P(A|B)^c P(B^c)}$$

Example2:

- With reference to Example 1, if someone is considering joining the club but delays his decision for several weeks only to find out that the fees have been increased, what is the probability that Ms. Catral was elected president of the club?

Solution: Using Bayes' theorem to write

$$P(B_3/A) = \frac{P(B_3)P(A/B_3)}{P(B_1)P(A/B_1) + P(B_2)P(A/B_2) + P(B_3)P(A/B_3)}$$

Continuation of Solution:

- And then substituting the probabilities calculated in Example 1, we have

$$P(B3/A) = \frac{0.8}{0.24 + 0.05 + 0.08} = 8/37$$

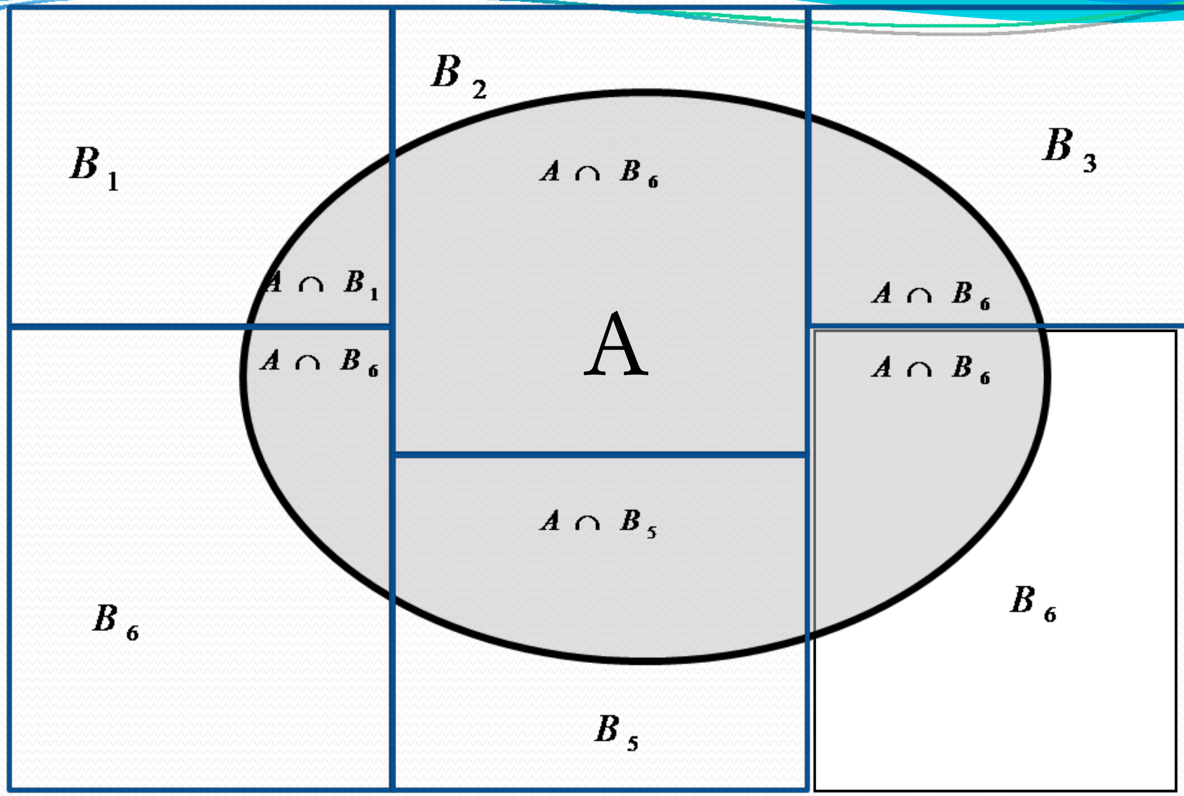
In view of the fact that fees have increased, this result suggests that Ms. Catral is probably not the president of the club.

Notes on Theorem 10.10 & 10.11

- Both theorems require the same conditions that

  $\{B_1, B_2, \ldots, B_n\}$ is a collection of mutually exclusive events wherein each event has a nonzero probability and $\Omega = B_1 \cup B_2 \cup \ldots \cup B_n$. These conditions imply that each time we perform the experiment, exactly one of the events $B_1, B_2, \ldots, B_n$ will occur.

  We can visualize this using the Venn Diagram shown in the next page (Figure 10.4).



- FIGURE 10.4. VENN DIAGRAM WHERE EVENTS B_1, B_2, B_3, B_4, B_5, B_6 ARE MUTUALLY EXCLUSIVE EVENTS AND $B_1 \cup B_2 \cup \ldots \cup B_6 = \Omega$. Usually, the B_i's are DEFINED IN TERMS OF THE FIRST STAGE OF THE EXPERIMENT WHILE EVENT A IS DEFINED IN TERMS OF THE SECOND STAGE OF THE EXPERIMENT.

Notes:

- The theorem of total probability and Bayes' theorem are useful when there are two stages in an experiment.
- Oftentimes, we specify the events B_1, B_2, …, B_n in terms of the first stage of the experiment while we specify event A in terms of the second stage of the experiment.
- We can then view both theorems as a more natural way of computing probabilities because the formulas tell us that we first need to consider all the possible ways in which the first stage of the experiment can occur in order to facilitate the computation of the probability of an event described in terms of the second stage of the experiment.

Biostatistics

- Statistics is very useful in analyzing biological or medical data. This area of statistics is called *biostatistics*.

- One special interest in biostatistics is to determine the effectiveness of a medical test in detecting a particular disease.

- The *sensitivity* of a given test is the probability of correctly diagnosing a person who has the disease while the *specificity* of a test is the probability of correctly diagnosing a person who does not have the disease. Naturally, we would want both the sensitivity and specificity of a test to be very close to 1.

Read Example 10.30:

- Suppose that it is known that only 0.005 of the people in a town have diabetes. A diabetes test is available. The sensitivity of this test is 0.999 and its specificity is 0.995. Suppose a person from this town is selected at random and the diabetes test was performed.

a) What is the probability that the test will indicate that he has diabetes?

b) If the test shows that the person has diabetes, what is the probability that the test is correct, that is, he really does have diabetes?

c) If the test shows that the person does not have diabetes, what is the probability that the test is correct, that is, he really does not have diabetes?

Solution to Example 10.30

- We can view the experiment involved in this probability problem as being performed in two stages.
- The first stage involves the selection of the person and observing whether or not he has diabetes.
- The second stage involves testing the selected person and observing the outcome of the test.

Let A = event that the test of the selected person shows he has diabetes

B = event that the selected person actually has diabetes

Con't of Solution to Example 10.30

- Given: $P(B) = 0.005$. $P(B^c) = 1 - P(B) = 0.995$.
- We can express both the sensitivity of the test and specificity of the test as conditional probabilities.
- The sensitivity of the test is given and we can express it as $P(A|B)=0.999$.
- The specificity of test is also given and we can express it as $P(A^c|B^c)=0.995$.

Con’t of Solution to Ex. 10.30

- Using the property of the conditional probability as a probability function, we can compute:

$P(A^c|B)=1 - P(A|B) = 0.001$

$P(A|B^c) = 1 - P(A^c|B^c) = 0.005.$

Find P(A). We solve this by using the corollary of the Theorem of Total Probability.

$$P(A) = P(A|B)P(B) + P(A|B^c)P(B^c)$$
$$= (0.999)(0.005) + (0.005)(0.995) = 0.00997$$

Con't of Solution to Ex. 10.30

b) Find P(B|A). We solve this by using the corollary of Bayes' Theorem.

$$P(B|A)=\frac{P(A|B)P(B)}{P(A|B)P(B)+P(A|B^c)P(B^c)}=\frac{(0.999)(0.005)}{0.00997}=0.5010$$

c) Find $P(B^c|A^c)$. We also solve this by using the corollary of Bayes' Theorem.

$$P(B^c|A^c)=\frac{P(A^c|B^c)P(B^c)}{P(A^c|B)P(B)+P(A^c|B^c)P(B^c)}=\frac{(0.995)(0.995)}{(0.001)(0.005)+(.995)(.995)}=0.999995$$

Note on Example 10.30

- We can be more confident about the result of the test if the result of the test were negative than if the result of the test were positive.
- This is because the probability is already very close to 1 that a person really does not have diabetes if the result of the test were negative as compared to only a 50-50 chance that a person really has diabetes if the result of the test were positive.

- Homework: Do exercises on pp. 323 to 324 (exercises for section 10.6)

10.7 Independent Events

Definition 10.17: Independent events

Two events A and B are said to be **independent events** if and only if any one of the following conditions is satisfied:

a) $P(A|B) = P(A)$ if $P(B)>0$; or,

b) $P(B|A) = P(B)$ if $P(A)>0$; or,

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

Otherwise, the events are said to be **dependent.**

Notes on Independent Events

- The three stated conditions in the definition are logically equivalent to each other.
- This means that if we know that one of these three conditions is true then we are assured that the other two conditions will also be true.
- Likewise, if we know that one of the three conditions is false then we are also assured that the other two conditions will also be false.
- This is why it is sufficient to check the veracity of just one of the three stated conditions in the definition to determine whether two events are independent or not.

Example for independent events

1. Consider the following events in the toss of a single die:

   A: Observe an odd number

   B: Observe an even number

   Are A and B independent events?

2. The probability that Robert will correctly answer the toughest question in an exam is ¼. The probability that Ana will correctly answer the same question is 4/5. Find the probability that both will answer the question correctly, assuming that they do not copy from each other.

Example 10.31:

- Consider the experiment of tossing a fair die twice.

  Define

  A=event of an even number of dots on the first toss

  B=event of observing more than 4 dots on the second toss
  C=event of observing less than 6 dots on the first toss

- Using the classical definition of probability, we get the following probabilities:

  P(A)=18/36=1/2 P(B)=12/36=1/3 P(C)=30/36=5/6

Example 10.31:

- Working on the reduced sample space, we get the following conditional probabilities:

  P(A|B)=6/12=1/2   P(A|C)=12/30=2/5
  P(B|C)=10/30=1/3

  We can then conclude that A and B are independent events because the first condition, P(A)=P(A|B), is satisfied.

Example 10.31:

- We can easily verify that the second and third conditions in the definition are both true.

  P(B) is indeed equal to P(B|A).

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

- Events B and C are also independent events since P(B)=P(B|C).
- However, A and C are dependent events because the first condition is not satisfied, that is, $P(A) \neq P(A|C)$. The computed results indicate that if we know event C occurred then our assigned probability for event A will decrease from the original measure of 1/2 to the new measure of 2/5.

Notes on Independent Events

- By examining the three conditions in the definition closely, we will notice that the first two stated conditions are consistent with the layman's concept of independence.

- According to these statements, we say that two events A and B are independent of each other whenever the conditional probability of A given B is just the same as the unconditional probability of A, or, the conditional probability of B given A is just the same as the unconditional probability of B.

Notes on Independent Events

- This is just the same as saying that the occurrence of event B does not affect the assigned probability for event A, or, the occurrence of event A does not affect the assigned probability for event B.

- While the first two conditions stated in the definition give us a clear picture of the concept of independence, it is the third condition that helps us understand the importance of independence.

Notes on Independent Events

- As stated in the third condition, we only need to know the individual probabilities of two events in order to compute the probability that these two events will occur simultaneously.
- In other words, satisfying the condition of independence of two events facilitates the computation of the probability that these events will occur simultaneously.

Example 10.32

Suppose A and B are independent events with $P(A)=0.3$ and $P(B)=0.6$. It is easy to compute for the probabilities of the following events:

a) $P(A \cap B) = P(A)P(B) = (0.3)(0.6) = 0.18$

b) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

$= 0.3 + 0.6 - (0.3)(0.6) = 0.72$

c) $P(A \cap B^c) = P(A) - P(A \cap B) = 0.3 - (0.3)(0.6) = 0.12$

Notes on Example 10.32

- Many students confuse independent events with mutually exclusive events.
- This is probably because both concepts facilitate the computation of probabilities.
- It should be clear though that these two concepts, though related, are actually distinct properties.
- If A and B are mutually exclusive events then $A \cap B = \varnothing$ so that $P(A \cap B) = 0$.
- Whereas, if A and B are independent events then, by definition, $P(A \cap B) = P(A)P(B)$.

Notes on Example 10.32

- As a result, events A and B will be mutually exclusive and independent events whenever $P(A \cap B) = P(A)P(B) = 0$.
- For this condition to be true, we require that at least one of events A or B has zero probability.
- In fact, if A and B are mutually exclusive events and both have nonzero probabilities then it is impossible for them to be independent events at the same time.
- Likewise, if A and B are independent events and both have nonzero probabilities then it is impossible for them to be mutually exclusive.

Example 10.33

Suppose A and B are events with $P(A)=0.4$ and $P(B)=0.3$. Determine $P(A \cap B)$ and $P(A \cup B)$ based on the given assumption.

a) Under the assumption that A and B are independent events:

Solution to Example 10.33a)

$P(A \cap B) = (0.4)(0.3) = 0.12$ (by definition of independence)

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$

$= 0.4 + 0.3 - 0.12 = 0.58$

We will notice that $P(A \cap B) \neq 0$. This leads us to the conclusion that $A \cap B \neq \varnothing$; that is, A and B are not mutually exclusive events.

Solution to Example 10.33 b)

b) Under the assumption that A and B are mutually exclusive events:

$P(A \cap B) = 0$ (since $A \cap B = \emptyset$)

$P(A \cup B) = P(A) + P(B) = 0.4 + 0.3 = 0.7$

(by finite additivity)

We will notice that $P(A \cap B) \neq P(A) x P(B)$; that is, A and B are not independent events.

Notes on Example 10.33:

- An obvious consequence of the independence of events A and B is the independence of the events A and B^c.
- This is the same as saying that if $P(A)=P(A|B)$ then $P(A)=P(A|B^c)$.
- This implies that if A and B are independent events then our assigned probability for event A will not change even if we have information on event B as regards to whether it occurred or not, since $P(A)=P(A|B)=P(A|B^c)$ under the assumption of independence.

Notes on Example 10.33:

- Aside from events A and B^c, there are other pairs of independent events whenever A and B are independent events.

- These are events A^c and B; as well as, events A^c and B^c.

- We leave the proof of these results as exercises at the end of this section.

- Homework: Do pp. 326 to 327 (Exercises for section 10.7)

10.8 Random Variables and Distributions

Definition 10.18: Random Variable

A function whose value is a real number that is determined by each sample point in the sample space is called a **random variable.**

An uppercase letter, say X, will be used to denote a random variable and its corresponding lowercase letter, x in this case, will be used to denote one of its values

Notes on Random Variable

- The addition of the term "random" emphasizes the requirement that the realized or actual value of the random variable depends on the outcome of a random experiment.

- Consequently, it is impossible to predict with certainty what the realized value of the random variable X will be.

Notes on Random Variable

- Since the random variable is a function, as stated in the definition, then each outcome in the sample space must be mapped to exactly one real number.
- No sample point can be assigned more than one real number; nor, can there be a sample point that is not assigned to any real number.
- This assures us that the random variable X will have one and only one realized value, whatever the outcome of the random experiment.

Example 1

- An experiment consists of tossing a coin 3 times and observing the result. What are the possible outcomes and the values of the random variables X and Y, where X is the number of heads and Y is the number of heads minus the number of tails?

Example 2

- A hatcheck girl returns 3 hats at random to 3 customers who had previously checked them. If Jason, Charlie, and Ohmar, in that order, receives one of the hats, list the sample points for the possible orders of returning the hats and find the values m of the random variable M, that represents the number of correct matches.

Read Example 10.34

- Filipinos are so fascinated with elections and the polls conducted to predict the outcomes of these elections.
- For illustration purposes, let us imagine a very small barangay consisting of 6 qualified voters.
- Let's label these voters as A1, A2, A3, A4, A5, and A6. There are two candidates vying for the position, say Renzo and Sandro.

Example 10.34

- What we do not know is that voters A1, A2, A3 and A4 have already decided to elect Renzo while voters A5 and A6 will elect Sandro.
- We only have enough resources to get a sample of size 3.
- We will then use the information from this sample to predict the outcome of the election.
-

Example 10.34

- Suppose we use simple random sampling without replacement to select our sample of size 3. Our sample space will contain all the 20 possible combinations of size 3. The sample points in our sample space are:
- {A1,A2,A3} {A1,A2,A4} {A1,A2,A5} {A1,A2,A6}
  {A1,A3,A4} {A1,A3,A5} {A1,A3,A6} {A1,A4,A5}
  {A1,A4,A6} {A1,A5,A6} {A2,A3,A4} {A2,A3,A5}
  {A2,A3,A6} {A2,A4,A5} {A2,A4,A6} {A2,A5,A6}
  {A3,A4,A5} {A3,A4,A6} {A3,A5,A6} {A4,A5,A6}

Example 10.34

- Define X=number of voters who will elect Renzo.

  X is a random variable because we can map each one of the enumerated sample point into one and only one real number.

  For instance, if the selected sample were {A1,A2,A3} then the realized value of X is 3 because all 3 voters in the sample will elect Renzo; while if the selected sample were {A4,A5,A6} the realized value of X is 1 because only A4 will elect Renzo.

Example 10.34

- The realized values of X for the other sample points in the sample space are as follows:

Sample Point	x	Sample Point	x	Sample Point	x	Sample Point	x
{A1,A2,A3}	3	{A1,A2,A4}	3	{A1,A2,A5}	2	{A1,A2,A6}	2
{A1,A3,A4}	3	{A1,A3,A5}	2	{A1,A3,A6}	2	{A1,A4,A5}	2
{A1,A4,A6}	2	{A1,A5,A6}	1	{A2,A3,A4}	3	{A2,A3,A5}	2
{A2,A3,A6}	2	{A2,A4,A5}	2	{A2,A4,A6}	2	{A2,A5,A6}	1
{A3,A4,A5}	2	{A3,A4,A6}	2	{A3,A5,A6}	1	{A4,A5,A6}	1

- Clearly, the realized value of X depends on which sample of size 3 we actually select for our study to predict the outcome of the election

Notes on Random Variables

- The concept of a random variable will provide us with a new way of expressing events.
- In particular, we will use the notation, $X \leq x$, to express the event containing all sample points whose associated value for the random variable X is less than or equal to x, where x is a specified real number.
- We will use the notation, $X > x$, to express the event containing all sample points whose associated value for X is greater than x.

Notes on Random Variables

- We will use the notation, $a<X<b$, to express the event containing all sample points whose associated value for X is in between a and b, where a and b are specified real numbers. And so on.

- Read Example 10.35 – this provides us with illustrations on how we use the random variable to express the events whose probabilities we are interested in

Read Example 10.35

Let us use the random variable defined in Example 10.34, X=number of voters who will elect Renzo. Specify the following events by roster method and express it in terms of X:

- A = event of selecting a sample with 1 voter electing Renzo
- B= event of selecting a sample with more than 2 voters electing Renzo
- C= event of selecting a sample with at least 1 voter electing Renzo
- D= event of selecting a sample with 5 voters electing Renzo

Solution to Exercise 10.35

- A={{A1,A5,A6}, {A2,A5,A6},{A3,A5,A6}, {A4,A5,A6}}

  We can express event A in terms of the random variable, X, as X=1 because this contains all of the sample points in Ω whose value for X is equal to 1.

- B ={{A1,A2,A3}, {A1,A2,A4}, {A1,A3,A4}, {A2,A3,A4}}

  We can express event B in terms of the random variable, X, as X>2 because this contains all of the sample points in Ω whose value for X is greater than 2.

Solution to Exercise 10.35

- $C = \Omega$ = sure event.

  We can express event C as $X \geq 1$.

- $D = \varnothing$ = impossible event.

  We can express event D as $X = 5$.

Notes on Example 10.35

- It is important to note that we used capital Latin letters to denote events and random variables, as well.

- Although we are using the same notation for events and random variables, we have to remember that these are actually two distinct though related concepts.

Notes on Example 10.35

- In Example 10.35, we used X to denote a random variable, while we used A, B, C and D to denote the events themselves.
- The random variable, X, by itself is not an event. The events in Example 10.35 are $X=1$, $X>2$, $X\geq 1$ and $X=5$.

Notes on Example 10.35:

- Let us not forget that we will view the characteristic of interest in a particular study as a random variable whose value depends on the outcome of a random experiment.

- In many studies, researchers aim to come up with a probabilistic/stochastic model that they can use to describe the essential properties of this random variable. One tool that researchers use in modeling is the *cumulative distribution function* of the random variable

Definition of CDF

Definition 10.19

- The **cumulative distribution function** (cdf) of a random variable X, denoted by F(.) is a function defined for any real number x as

$$F(x) = P(X \le x)$$

Notes on the CDF

- Since the CDF is defined in terms of a probability function then its value will range from 0 to 1, inclusive of endpoints

- It is a non-decreasing function. This means that it is impossible for the value of the CDF, F(x), to decrease as the value of x increases.

- Thus, the graph of the CDF of a random variable in any interval of the real line either remains flat or it goes up, but it will never go down.

Notes on the CDF

- Every random variable will have one and only one CDF.
- We can use the CDF to compute for the probability of any event that is expressed in terms of the random variable.
- Example: X is the IQ of a selected person from the population and we know the CDF of X then we can determine the probability of selecting a person whose IQ is below average or the probability of selecting a person whose IQ is greater than 100.
- The behavior of the CDF of a random variable depends on the type of random variable.

4 types of random variables

- *discrete random variables*
- *continuous random variables*
- *singular continuous random variables*
- *mixed random variables.*

Definition of Discrete Sample Space and Discrete Random Variable

Definition 10.20

- If a sample space contains a finite number of sample points or has as many sample points as there are counting/natural numbers then it is called a **discrete sample space**.

Definition 10.21

- A random variable defined over a discrete sample space is called a **discrete random variable**.

Notes on Discrete Sample Space and Discrete Random Variable

- The sample space used to define a discrete random variable either has a finite number of sample points, or, has as many sample points as there are counting numbers.

- The discrete random variable can contain infinitely many sample points so long as this collection has a one-to-one correspondence with the collection of counting numbers.

Read Example 10.36

- The random variable defined in Example 10.34, X=number of voters who will elect Renzo is an example of a discrete random variable.
- This is because the sample space used to define it contains a finite number of sample points.
- Remember that the sample space that we specified in Example 10.34 contains 20 sample points.

Read Example 10.37 - Example of Discrete Sample Space:

Example 10.37: Consider the experiment of tossing a coin until a head comes up. Define its sample space as follows:

Ω = {H, TH, TTH, TTTH, TTTTH, …}

- Notice that this sample space contains infinitely many sample points but it has a one-to-one correspondence with the set of counting or natural numbers.
- To show this, we can define the correspondence by counting the number of tosses required to perform the experiment. By doing so, we will be able to pair the outcome H with 1 because this outcome required only 1 toss. Then, we will pair TH with 2, TTH with 3, and so on. We see that we are able to match each outcome with a unique counting number, and vice versa. We can then conclude that Ω has as many sample points as there are counting numbers.

Example of Discrete Random Variable

- Random variables defined on this sample space are discrete random variables.
- Examples of such random variables are X=number of tails and Y=number of tosses. The values of X and Y for the first 5 sample points will then be as follows:

Sample Point	x	y
H	0	1
TH	1	2
TTH	2	3
TTTH	3	4
TTTTH	4	5

- There are many subsets of the set of real numbers that have a one-to-one correspondence with the set of counting numbers.
- Examples of discrete sample space:
  - the set of all positive integers
  - the set of non-negative integers
  - the set of all integers
  - the set of even numbers
  - the set of odd numbers
  - the set of rational numbers.

Question: Is it always possible to find a one-to-one correspondence between any infinite set and the set of counting numbers?

- Answer: NO!
- Example, the set of all real numbers does not have a one-to-one correspondence with the set of counting numbers.
- All closed, open, half-open and unbounded intervals of real numbers such as $[0,1]$, $(0.2, 0.5)$, $(5, 100]$, $(1, \infty)$ do not have a one-to-one correspondence with the set of counting numbers.

Probability Mass Function

Definition 10.22

- The **probability mass function** (PMF) of a discrete random variable, denoted by f(.), is a function defined for any real number x as:

$$f(x) = P(X = x).$$

- The values of the discrete random variable X for which $f(x)>0$ are called its **mass points**.

Notes on PMF

- Just like the CDF, the probability mass function of a discrete random variable X can be used to compute for the probability of any event that is expressed in terms of X.
-
- We can also use the PMF to compute important summary measures such as the mean and the standard deviation.

Steps in Constructing the PMF of a discrete random variable X (By classical defn of probability to compute for P(X=x)

Step 1. Identify the mass points of X. The mass points of X are actually the possible values that X could take on because these are the points where P(X=x) will be nonzero. In other words, the set of mass points of X is the range of the function, X.

Step 2. Determine the event associated with the expression, X=x.

Step 3. Compute for the probability of this event.

Recall Example 1: Toss a coin 3 times

- Construct the probability mass function of the random variables X and Y

Recall Example 2: Hatcheck girl

- Construct the probability mass function for the random variable M.

Read Example 10.38

Let X=number of voters who will elect

Renzo, as defined in Example 10.34.

- The experiment involves the selection of a sample of size 3 using simple random sampling without replacement.
- This means that we will be selecting the sample in such a way that all of the 20 possible samples will be given the same chances of selection.
- We can then use the classical definition of probability because the outcomes in the sample space are equiprobable.
-

Event associated with X=x

x	
1	{{A1,A5,A6}, {A2,A5,A6},{A3,A5,A6}, {A4,A5,A6}}
2	{{A1,A2,A5}, {A1,A2,A6}, {A1,A3,A5}, {A1,A3,A6}, {A1,A4,A5}, {A1,A4,A6}, {A2,A3,A5}, {A2,A3,A6},{A2,A4,A5}, {A2,A4,A6}, {A3,A4,A5}, {A3,A4,A6}}
3	{{A1,A2,A3}, {A1,A2,A4}, {A1,A3,A4}, {A2,A3,A4}}

- Using the classical definition of probability, we can compute:
- $f(1) = P(X=1) = P(\{\{A1,A5,A6\}, \{A2,A5,A6\},\{A3,A5,A6\}, \{A4,A5,A6\}\} = 4/20=1/5.$
- $f(2)=P(X=2)=12/20=3/5.$
- $f(3)=P(X=3) = 4/20=1/5.$

- The probability mass function of X is:

$$f(x)=\begin{cases} 1/5 & when\ x=1 \\ 3/5 & when\ x=2 \\ 1/5 & when\ x=3 \\ 0 & for\ all\ other\ real\ numbers \end{cases}$$

PMF in table form

X	*1*	*2*	*3*
f(x)	1/5	3/5	1/5

- The sum of the pmf must always be equal to 1.

Using pmf to determine the probability of an event

- If we want to evaluate P(a < X < b) then we can easily do this using the pmf by following these steps:
- *Step* 1. Identify the mass points, x, that are included in the interval of interest.
- *Step* 2. Use the pmf to determine the value of P(X=x) for each one of the mass points identified in Step 1.
- *Step* 3. Get the sum of all the values derived in Step 2.

Example 1: toss a coin 3 times

Find the following:

a) $P(X < 3)$

b) $P(1 \leq X \leq 3)$

c) $P(X > 1)$

Read Example 10.39 - Example of Using pmf for Computing Probabilities

Example 10.39: Use the pmf that we derived in Example 10.38 to determine the following probabilities:

a) $P(X<3)$ b) $P(2 \leq X \leq 5)$ c) $P(X>2)$

Solution:

a) $P(X<3) = P(X=1) + P(X=2) = 4/5$.

b) $P(2 \leq X \leq 5) = P(X=2) + P(X=3) = 4/5$.

c) $P(X>2) = P(X=3) = 1/5$.

Probability Density Function

Definition 10.23

- The **probability density function** (pdf) of a continuous random variable X, denoted by f(.), is a function that is defined for any real number x and satisfy the following properties:
- a) $f(x) \geq 0$ for all x;
- b) the area below the whole curve, f(x), and above the x-axis is always equal to 1; and,
- c) $P(a \leq X \leq b)$ is the area bounded by the curve f(x), the x-axis and the lines x=a and x=b.

- For a continuous random variable, the $P(X=a)=0$ for any real number, a.

- Thus, $P(X\leq a) = P(X<a) + P(X=a) = P(X<a)$. For the same reason, we can also say that whenever X is a continuous random variable:

$$P(a < X < b)=P(a \leq X < b)=P(a < X \leq b)=P(a \leq X \leq b).$$

Area of shaded region

- Rectangle = length x width

Example of PDF: skip

- A continuous random variable X that can assume values between 0 and 2 has a density function given by

$$f(x) = \begin{cases} 0.5 & \text{for } 0 < x < 2 \\ 0 & \text{otherwise} \end{cases}$$

Find the following probabilities:

a. $P(1 < X < 2)$
b. $P(X > 1.5)$

Skip

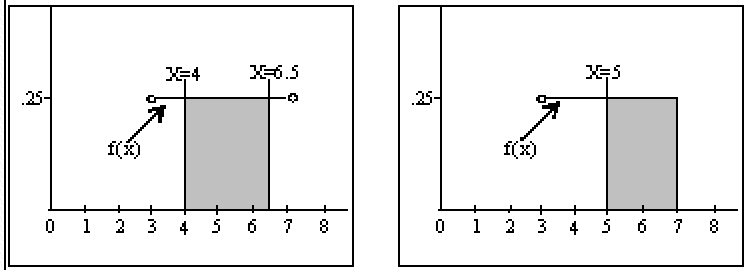
- **Example 10.40:** The pdf of a continuous random variable, X, is given by
- Find the following probabilities:
- a) $P(4 < X < 6.5)$
  b) $P(X \geq 5)$

Solution: Figure 10.7 shows the graph of the pdf of the random variable X and the respective regions of interest. Since the shaded regions are always in the form of a rectangle, we will use the formula for the area of a rectangle given by: Area = Length x Width.

- a) Length=0.25 and Width=2.5

  $$P(4 < X < 6.5) = (0.25)(2.5) = 0.625$$

- b) Length = 0.25 and Width=2



a) P(4 < X < 6.5)

b) P(X ≥ 5)

- Do exercises for Section 10.8 (nos. 1, 3, and 5)

10.9 Expected value of a Random Variable

Definition 10.24

Let X be a discrete random variable with probability mass function:

x	x_1	x_2	…	x_n
f(x)=P(X=x)	$f(x_1)$	$f(x_2)$	…	$f(x_n)$

The **expected value of X**, also referred to as the **mean of X** is:

$$E(X) = \mu = x_1 f(x_1) + x_2 f(x_2) + \ldots + x_n f(x_n) = \sum x_i f(x_i)$$

Examples of Expected Values

1. Find the mean of the random variables X and Y of Experiment 1 (Toss of a coin 3 times).

2. Find the expected number of correct matches in Experiment 2 (Hatcheck girl).

3. In a gambling game, a man is paid P50 if he gets all heads or all tails when 3 coins are tossed, and he pays out P30 if either 1 or 2 heads show. What is his expected gain?

Theorem 10.12

Let X be a discrete random variable with probability mass function given by:

x	x_1	x_2	…	x_n
$f(x)=P(X=x)$	$f(x_1)$	$f(x_2)$	…	$f(x_n)$

Suppose Y=g(X) is a discrete random variable then the **expected value of g(X)** is

$$E(g(X)) = \sum g(x_i)f(x_i)$$

Example 10.44:

A used car dealer finds that in any day, the probability of selling no car is 0.4, one car is 0.2, two cars is 0.15, 3 cars is 0.10, 4 cars is 0.08, five cars is 0.06 and six cars is 0.01. Let X=number of cars sold and let Y=500+1500X represent the salesman's daily earnings. Find the salesman's expected daily earnings.

Definition 10.25

Let X be a random variable with mean, μ. The **variance of X**, denoted by σ^2 or Var(X), is defined as $\sigma^2 = Var(X) = E(X - \mu)^2$. The **standard deviation of X** is the positive square root of the variance.

Definition 10.26

Let X be a discrete random variable with pmf given by:

x	x_1	x_2	…	x_n
$f(x)=P(X=x)$	$f(x_1)$	$f(x_2)$	…	$f(x_n)$

The variance of X is:

$$\sigma^2 = Var(X) = E(X - \mu)^2 = \sum (x_i-\mu)^2 f(x_i)$$

Theorem 10.13

If X is a random variable then the $Var(X) = E(X^2) - \mu^2$.

Example:

In experiment number 1, find the variance of X using definitional and computation formulas.

Properties of the Mean and Variance

Let X and Y be random variables (discrete or continuous) and let a and b be constants.

1. $E(aX + b) = aE(X) + b$

   Special Cases: a. If $b = 0$, then $E(aX) = aE(X)$

   b. If $a = 0$, then $E(b) = b$

2. $E(X + Y) = E(X) + E(Y)$

   $E(X - Y) = E(X) - E(Y)$

3. $E(XY) = E(X)E(Y)$ if X and Y are independent

4. $E[X - E(X)] = 0$

Properties of the Mean and Variance

Let X and Y be random variables (discrete or continuous) and let a and b be constants.

5. $Var(aX + b) = a^2Var(X)$

   Special Cases:

   a) if $b = 0$, then $Var(aX) = a^2Var(X)$

   b) if $a = 0$, then $Var(b) = 0$

6. If X and Y are independent then

   $Var(X + Y) = Var(X) + Var(Y)$

   $Var(X - Y) = Var(X) + Var(Y)$

Example: Properties of Mean and Variance

If X and Y are independent random variables with $E(X) = 3$, $E(Y) = 2$, $Var(X) = 2$ and $Var(Y) = 1$, find

a) $E(3X + 5)$
b) $Var(3X + 5)$
c) $E(XY)$
d) $Var(3X-2Y)$

Some Notes:

We also use the concept of expectation to define measures of skewness and kurtosis of the pmf of a discrete random variable and the pdf of a continuous random variable.

The degree of asymmetry of a distribution is often measured using the coefficient of skewness, μ_3/σ^3, where $\mu_3 = E(X - \mu)^3$. Its interpretation is the same as other measures of skewness. In general,

$\mu_3/\sigma^3 < 0$ – skewed to the left

$\mu_3/\sigma^3 > 0$ – skewed to the right

$\mu_3/\sigma^3 = 0$ – symmetric

Some Notes:

The “peakedness” of the distribution, on the other hand, is often measured by the coefficient of kurtosis, μ_4/σ^4, where μ_4=E(X - μ)^4. In general,

$\mu_4/\sigma^4 < 3$ – platykurtic

$\mu_4/\sigma^4 > 3$ – leptokurtic

$\mu_4/\sigma^4 = 3$ – mesokurtic

Homework

- Do exercises for Section 10.9, nos. 1, 2 , 3 and 5.