Stat 121 (Mathematical Statistics I)

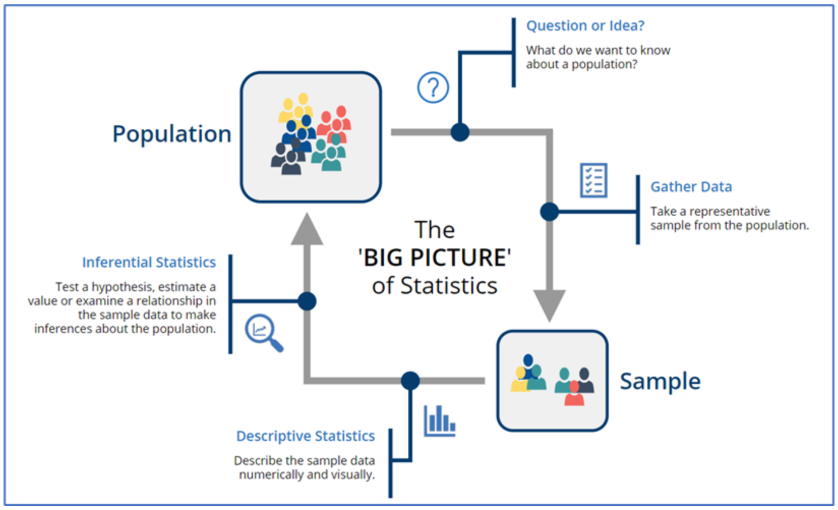
Lesson 1.1: Properties of Probability

Learning Outcomes

1. **Explain** what is probability,
2. **Differentiate** the ways of assigning probabilities to events, and
3. **Compute** probability of events.

Introduction

Statistics is the art and science of using sample data to understand something about a population. In Figure 1 we illustrate the process of doing a statistical inquiry. In statistics, we always start with a question or hypothesis about the characteristic (parameter) of a population. We wanted to collect data for us to be able to provide an answer to this question or to test our hypothesis. However, populations are generally large such that collecting information from each unit in the population would be costly and time-consuming. Hence, what we usually do is randomly select a subset (sample) of elements from the population and gather the required data from the sample. The summary of the information (statistics) obtained from the sample are then used to provide an answer to the question about the population. It could be in the form of an estimate or a test statistic.



Since, the sample is relatively very small compared to the size of the population, there is always uncertainty attached to the estimate or the test statistic based on the data from the sample. This uncertainty is quantified by computing a number which we call as **probability** which represents the likelihood of an event occurring. With the use of probability, we can come up with a degree of confidence with our sample estimates. Probability and its properties are the main focus of this course. Before we discuss the details on what really is probability, we need to study first basic concepts that are important for us to get a better understanding of probability.

Random Experiment, Sample Space, and Event

A **random (probabilistic) experiment** is a process which can be repeated under similar conditions but whose outcome cannot be predicted with certainty beforehand. The list or set of all possible outcomes of a random experiment is called the **outcome space** or more popularly the **sample space**. The sample space being a set is denoted by a capital letter, commonly by S or Ω. Each element of the sample space is called a **sample point**, or simply an **outcome**.

There are many examples of random experiments ranging mostly from games of chance such as rolling a pair of dice, tossing coins, and drawing cards from a deck of playing cards. The reason for this is that the historical development of probability theory can be traced back to gambling in the early 1650s. You can find a lot of articles in Google about the history of probability.

Other examples of random experiments are: recording the weight of new-born infants, recording the score of students in a 100-point examination, observing the effectiveness of a COVID-19 vaccine, and many more.

Example 1.1

1. Suppose we roll a six-sided die and record the number of dots on the up face, the list of possible outcomes is $S = \{1, 2, , 3, 4, 5, 6\}$.
2. If we toss two coins, then the sample space is given by $S = \{HH, HT, TH, TT\}$, where H and T denotes a head and a tail, respectively. The outcome HH means that both coins land heads, an outcome HT means the first coin land head and the other a tail, and so on.
3. Suppose we randomly select a student, and ask them "how many pairs of jeans do you own?". In this case our sample space is $S = \{0, 1, 2, 3, ...\}$

Probability is calculated on a given event. **Events** are subsets of the sample space. It is a collection of outcomes. Since an event is a set, it is denoted by a capital letter, except letter S.

Example 1.2

1. Let A = event of observing odd number of dots in tossing a die. Then $A = \{1, 3, 5\}$.
2. Let B = event of observing at least one head in tossing a pair of coins. Then $B = \{HT, TH, HH\}$.
3. Let C = event that a student has at most 5 pairs of jeans. Then $C = \{0, 1, 2, 3, 4, 5\}$.

There are two special types of events: **null** or **impossible** event and **sure** or **certain** event. The null event is an event which is impossible to occur and is denoted by ϕ or $\{\}$, while the sure event is the event that will definitely occur and is equivalent to the sample space S.

Since events are sets, set operations such as *union*, *intersection*, and *complementation*, and their properties apply to events. Thus, you need to review the chapter on sets and set operations in Math 111s.

What is probability?

Mathematically speaking, probability is a number between 0 and 1, inclusive. Inclusive means it is possible to have a probability value equal to 0 or 1. The null event is assigned a probability equal to 0, while the sure event is given a probability of 1. A probability closer to 0 means the event is "less likely" to occur and a probability closer to 1 means the event is "highly likely" to occur.

Now we know that a probability is number between 0 and 1, inclusive. How does an event get assigned a particular probability value? Well, there are three ways of doing so:

1. the personal opinion (subjective) approach
2. the relative frequency approach
3. the classical approach

The **subjective** approach is the simplest in practice, but it is the least reliable. You might think of it as the "whatever it is to you" approach. Below are some examples.

Example 1.3

1. "I think there is an 35% chance of rain today."
2. "I think there is a 40% chance that well-known pharmaceuticals can come up with an effective COVID-19 vaccine before 2020 ends."
3. "I think I have a 75% chance of passing this course."

The **relative frequency** approach involves three steps in order to determine $P(A)$, the probability of an event A:

1. Perform an experiment a large number of times, n.
2. Count the number of times the event A of interest occurs, call the number $N(A)$.
3. Then, the probability of event A equals:

$$P(A) = \frac{N(A)}{n}$$

Example 1.4

Suppose the teacher requested three (3) students to toss a fair coin a large number of times and estimate the probability of a head. The table below shows the results of their tosses.

Student	No. of tosses (n)	Number of heads, N(A)	P(A)
Carmelo	4040	2054	0.5084
Lebron	10000	5067	0.5067
Dwayne	24000	12012	0.5005

As you can see the probability of a head is not 0.5 (your intuitive guess), but close to it. In fact, as the number of tosses gets larger, the probability of a head gets closer to 0.5.

Example 1.5

Some trees in a forest were showing signs of disease. A random sample of 200 trees of various sizes was examined yielding the following results:

Type	Disease free	Doubtful	Diseased	TOTAL
Large	35	18	15	68
Medium	46	32	14	92
Small	24	8	8	40
TOTAL	105	58	37	200

1. What is the probability that one tree selected at random is large?
2. What is the probability that one tree selected at random is diseased?
3. What is the probability that one tree selected at random is both small and diseased?
4. What is the probability that one tree selected at random from the population of medium trees is doubtful of disease?

SOLUTIONS:

The third approach in assigning probabilities is the **classical** or **equally-likely** approach. This approach assumes that every outcome in the sample space is equally likely to occur. Hence, the probability of event A is

$$P(A) = \frac{N(A)}{N(S)}$$

where $N(A)$ is the number of outcomes in event A and $N(S)$ is the number of outcomes in the sample space.

Example 1.6

Suppose you draw one card at random from a standard deck of 52 cards. Recall that a standard deck of cards contains 13 face values (Ace, 2, 3, 4, 5, 6, 7, 8, 9, 10, Jack, Queen, and King) in 4 different suits (Clubs, Diamonds, Hearts, and Spades) for a total of 52 cards. Assume the cards were manufactured to ensure that each outcome is equally likely with a probability of $\frac{1}{52}$.

1. What is the probability of obtaining a card that is a 2, 3, or 7?
2. What is the probability that the card drawn at random is a 2 of hearts, 3 of diamonds, 8 of spades or king of clubs?

SOLUTIONS:

Axioms of Probability

Probability is a (real-valued) set function P that assigns to each event A in the sample space S a number $P(A)$, called the probability of the event A, such that the following hold:

1. The probability of any event A must be non-negative, that is, $P(A) \geq 0$.
2. The probability of the sample space is 1, that is, $P(S) = 1$.
3. Given mutually exclusive events $A_1, A_2, A_3, \cdots$ that is, where $A_i \cap A_j = \phi$, for $i \neq j$, then

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

From these three (3) axioms, we develop numerous rules which help us assign probability to events.

1. **Complement Rule**: $P(A^c) = 1 - P(A)$

PROOF:

We can write $S = A \cup A^c$. Since A and A^c are mutually exclusive, by Axiom 3, we have

$$P(S) = P(A) + P(A^c)$$

and by Axiom 2, $P(S) = 1$, thus,

$$1 = P(A) + P(A^c)$$

which after transposition yields

$$P(A^c) = 1 - P(A)$$

2. **Null Set Rule**: $P(\phi) = 0$.

PROOF:

This follows immediately from the complement rule with $A = \phi$ and $A^c = S$.

3. **Upper Bound Rule**: $P(A) \leq 1$

PROOF:

In the proof of the complement rule, we wrote $P(S) = P(A \cup A^c) = P(A) + P(A^c)$, which by Axiom 2 becomes $1 = P(A) + P(A^c)$. Note that by Axiom 1, $P(A^c) \geq 0$. So that if we remove the non-negative term from the equation, $P(A)$ must be less than or equal to 1.

4. **Monotonicity Rule**: Suppose A and B are events such that $A \subset B$. Then $P(A) \leq P(B)$.

PROOF:

Since $A \subset B$ it must be true that $B = A \cup (B \cap A^c)$. Note that $A \cap (B \cap A^c) = \phi$, in other words, A and $B \cap A^c$ are mutually exclusive. Thus, from Axiom 3, we have

$$P(B) = P(A) + P(B \cap A^c)$$

From Axiom 1, $P(B \cap A^c) \geq 0$, so that removing it from the above equation imply that

$$P(B) \geq P(A)$$

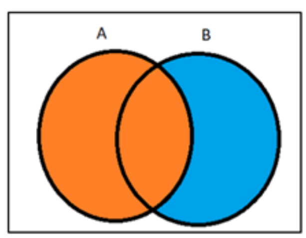
In other words,

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

5. **Addition rule**: For any two events A and B, $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

PROOF:

We write $A \cup B = A \cup (A^c \cap B)$. Refer to the diagram below.

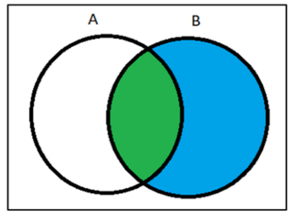


A is shaded with orange; $A^c \cap B$ is shaded with blue

Notice that A and $A^c \cap B$ are mutually exclusive, thus, by Axiom 3,

$$P(A \cup B) = P(A) + P(A^c \cap B) \tag{*}$$

Next we write, $B = (A \cap B) \cup (A^c \cap B)$. Refer to the diagram below.



$A \cap B$ is shaded with green; $A^c \cap B$ is shaded with blue

Observe also that $A \cap B$ and $A^c \cap B$ are mutually exclusive, thus, by Axiom 3,

$$P(B) = P(A \cap B) + P(A^c \cap B)$$

After transposition, we get

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

We then combine this with (*) to obtain the desired result

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

Example 1.7

A smoke detector system uses two interlinked units. If smoke is present, the probability the first unit will detect it is 0.95 and the probability the second unit will detect it is 0.90. The probability smoke will be detected by both units is 0.88.

a. If smoke is present, find the probability that the smoke will be detected by either unit or both.

b. Find the probability the smoke will go undetected.

SOLUTIONS:

Define the events $A = \{\text{first unit detects smoke}\}$ $B = \{\text{second unit detects smoke}\}$
We are given $P(A) = 0.95$, $P(B) = 0.90$, and $P(A \cap B) = 0.88$.

a. Using the addition rule, the probability that the smoke will be detected by either unit or both is given by

$$\begin{aligned} P(A \cup B) &= P(A) + P(B) - P(A \cap B) \\ &= 0.95 + 0.90 - 0.88 \\ &= 0.97 \end{aligned}$$

b. The smoke will go undetected if $A^c \cap B^c$. By DeMorgan's Law,

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

Hence,

$$\begin{aligned} P(A^c \cap B^c) &= P[(A \cup B)^c] \\ &= 1 - P(A \cup B) \\ &= 1 - 0.97 \\ &= 0.03 \end{aligned}$$

Remarks

1. The addition rule can be extended to more than 2 events. Suppose we have a finite sequence of events $A_1, A_2, A_3, \cdots, A_n$. Then

$$\begin{aligned} P(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i) - \sum_{i_1 < i_2} \sum P(A_{i_1} \cap A_{i_2}) \\ + \sum \sum_{i_1 < i_2 < i_3} \sum P(A_{i_1} \cap A_{i_2} \cap A_{i_3}) - \cdots + (-1)^{n+1} P(\bigcap_{i=1}^{n} A_i) \end{aligned}$$

For example, if $n = 3$, then

$$\begin{aligned} P(A_1 \cup A_2 \cup A_3) = P(A_1) + P(A_2) + P(A_3) - P(A_1 \cap A_2) \\ - P(A_1 \cap A_3) - P(A_2 \cap A_3) + P(A_1 \cap A_2 \cap A_3) \end{aligned}$$

2. **Bonferroni inequality**: $P(A \cap B) \geq 1 - P(A^c) - P(B^c)$

Example 1.8:

A company has bid on two large construction projects. The company president believes that the probability of winning the first contract is 0.6, the probability of winning the second contract is 0.4, and the probability of winning both contracts is 0.2.

a. What is the probability that the company wins at least one contract?

b. What is the probability that the company wins the first contract but not the second contract?

c. What is the probability that the company wins neither contract?

d. What is the probability that the company wins exactly one contract?

SOLUTIONS: Let $A =$ event of winning the first contract and $B =$ event of winning the second contract. We are given that $P(A) = 0.6$, $P(B) = 0.4$, and $P(A \cap B) = 0.2$.

a. The probability that the company wins at least one contract is

$$\begin{aligned} P(A \cup B) &= P(A) + P(B) - P(A \cap B) \\ &= 0.6 + 0.4 - 0.2 \\ &= 0.8 \end{aligned}$$

b. The probability that the company wins the first contract but not the second contract is given by

$$\begin{aligned} P(A \cap B^c) &= P(A) - P(A \cap B) \\ &= 0.6 - 0.2 \\ &= 0.4 \end{aligned}$$

c. The probability that the company wins neither contract is

$$\begin{aligned} P(A^c \cap B^c) &= P[(A \cup B)^c] \\ &= 1 - P(A \cup B) \\ &= 1 - 0.8 \\ &= 0.2 \end{aligned}$$

4. The probability that the company wins exactly one contract is

$$P(A \cup B) - P(A \cap B) = 0.8 - 0.2 = 0.6$$

Learning Tasks/Activities

Instruction: Answer the following as indicated.

1. Patients arriving at a hospital outpatient clinic can select one of three stations for service. Suppose that physicians are assigned randomly to the stations and that the patients therefore have no station preference. Three patients arrive at the clinic and their selection of stations is observed.

a. List the sample points for the experiment.

b. Let A be the event that each station receives a patient. List the sample points in A.

c. Make a reasonable assignment of probabilities to the sample points and find $P(A)$.

2. Consider an experiment in which each of three cars exiting from a university main entrance turns right (R) or left (L). Assume that a car will turn right or left with equal probability of 1/2.

a. What is the sample space S?

b. What is the probability that at least one car will turn left?

c. What is the probability that at most one car will turn left?

d. What is the probability that exactly two cars will turn left?

e. What is the probability that all three cars will turn in the same direction?

3. Find the probability of getting a numbered card when a card is drawn from the pack of 52 cards.

4. What is the probability of getting a sum of 7 or 9 when two dice are thrown?

5. During a visit to a primary care physician's office, the probability of having neither lab work nor referral to a specialist is 0.21. Of those coming to that office, the probability of having lab work is 0.41 and the probability of having a referral is 0.53. What is the probability of having both lab work and a referral?