1 Introduction

1.1 Why study probability and statistics?

- At the root of most scientific studies is the need to understand/measure some characteristics of a population of objects or of people. Measurements in a population usually exhibit some variation, large or small. In this course, we will see how we can take into account this variation in order to make intelligent decisions about the population.

- The population of interest is usually large and cannot be exhaustively studied. Instead, it is sampled in some intelligent/representative way. All calculations are made on the sample. We calculate the mean, the variance etc on the sample.

- From the sample, we wish to draw inference on the population. The meaningful and mathematically justifiable transition from the sample to the population is provided by the study of probability and statistics.

Example1 Coin tossing. A coin has two sides, head or tail. It is characterized by the likelihood of getting heads (or equivalently of getting tails). If we know it is perfectly balanced, then we may assume that the likelihood of getting heads is 50%. BUT we often do not have information about the coin. AND so we need to collect it. We do that by tossing the coin a few times in a fair way and by keeping track of how many heads come up. The results of the tosses is what constitute a sample. If we observe 100 heads in 100 tosses, can we then say with confidence that the coin is biased in favor of heads? What if we observe 65 heads? 50 heads? How does our level of confidence change each time?

Example2 Measuring the health of a population. We believe that the average weight of the population in the USA has increased since the last decade. We do not have the resources to weigh everyone. And so we take a representative "small" sample and weigh everyone in the sample. Suppose that a decade ago, the average weight for a sample of men was 130 pounds whereas today for a different sample it is 145 pounds. There is an observed difference of 15 pounds in the average. Can we then assert that the population average weight is on the rise?

Example3 We would like to measure the average lifetime of a new type of light bulb. The measurement of the lifetime of a single light bulb destroys the light bulb. Consequently, the need for collecting a sample is very evident.

Example4 Suppose that i phones are packed 25 in a box. Before shipping, there is a quality control process. Five phones are selected at random without replacement and given a thorough final check. If a single phone is found to be defective, the shipment is stopped. What are the merits of this sampling plan?

Example5 Suppose a pharmaceutical company claims they have a drug to cure the common cold. How can we test this conjecture? We may design a double blind experiment involving 20 individuals who are experiencing a cold. The individuals are chosen to be alike in many respects. Two groups of 10 each are formed. The first group all receive the new drug while those in the group will receive a placebo. Suppose that we observe 6 individuals in the treatment and 4 in the control group are cured. Can we conclude that the treatment is effective?

1.2 Discrete and continuous data

Discrete data is associated with the ability to enumerate or count the number of items in a set. We encounter discrete data when we count how many heads we observe in 100 tosses of a coin, or how often red occurs in spins of a roulette wheel, or how many electronic components are found defective in a box. We also have discrete data when we count how many tosses of a coin are required to observe the first occurrence of a head. It may take 1, 2,...etc tosses.

Continuous data are associated with an interval of values. We encounter continuous data when we "measure" something. For example, when we measure the height or weight of an individual we end up with continuous data. Volume, length and time are also continuous.

1.3 Probability

Definition The set of all possible outcomes of a statistical experiment is called the sample space and denoted by the symbol S. Elements in S will be denoted by small letters s.

Example We can display data using a tree diagram. Consider 3 tosses of a coin. We may write the set of possible outcomes in set notation,

$$S = \{(HHH), (HHT), (HTH), (HTT), (THH), (THT), (TTH), (TTT)\}$$

Example Suppose we have the experiment of choosing a point in the interval $(0, 1)$. The sample space could be denoted by

$$S = \{x : 0 < x < 1\}$$

Example Measuring the lifetime of an electrical component: $S = \{x : 0 < x < \infty\}$

Example Counting the number of calls received by an operator: $S = \{x : x = 1, 2, ...\}$

Definition An event is a subset of a sample space. It is denoted by a capital letter A,B,...Hence, the event

$$A = \{(HHH), (HHT), (HTH), (HTT)\}$$

is the event that H occurs in the first of three tosses of a coin.

We shall say that the event A above has occurred if any one of the possibilities

$$(HHH), (HHT), (HTH), (HTT)$$

has occurred. So if we observe (HTH) then we shall say that A has occurred.

Definitions Recall some set theory

- The intersection of two events A, B denoted $A \cap B$ is the event containing all elements common to both A and B.
- Two events A, B for which $A \cap B = \phi$, that is for which the intersection is the empty set are said to be disjoint. They have no element in common.
- The union of two events A, B denoted $A \cup B$ is the event containing all elements in both A and B.
- The complement of an event A with respect to the sample space S is the subset of all elements of S that are not in A. It is denoted by A'
- De Morgan's laws

$$
\begin{aligned}
(E_1 \cup ... \cup E_n)' &= E_1' \cap ... \cap E_n' \\
(E_1 \cap ... \cap E_n)' &= E_1' \cup ... \cup E_n'
\end{aligned}
$$

We use Venn diagrams often to illustrate events. This serves to simplify notation and serves to translate the English description of an event into set notation.

Example Let A, B, C be three events.

at least one of A,B,C	$A \cup B \cup C$
none of A,B,C	$A' \cap B' \cap C' = (A \cup B \cup C)'$
All three of A,B,C	$A \cap B \cap C$
exactly one of A,B,C	$(A \cap B' \cap C') \cup (A' \cap B \cap C') \cup (A' \cap B' \cap C)$
neither A nor B	$A' \cap B'$
exactly two of A,B,C	

1.4 Counting sample points

We will need to develop fast algorithms for counting the number of elements in a set. What is the number of possible outcomes when tossing a coin 5 times? How many possible hands of 5 cards each are there when chosen from a deck of 52 cards? How many possible hands of 13 cards each are there when chosen from a deck of 52 cards?

The multiplication principle states that if an experiment results in n_1 possible outcomes and a second experiment results in n_2 possible outcomes, then the performance of the first experiment followed by the second will result in $n_1 n_2$ possible outcomes.

Similarly for r experiments, the number of possible outcomes will be given by the product $n_1 n_2 ... n_r$.

Definition1 A permutation is an arrangement of all or part of a set of objects.

Theorem1 The number of permutations of n distinct objects is $n!$

Theorem2 The number of permutations of n objects, $n_1$of which are alike of one kind, $n_2$of which are alike of another kind, ...,$n_r$of which are alike of another kind is

$$\frac{n!}{n_1!...n_r!}$$

Example The number of permutations of the letters a, a, b, c, c, c, d is

$$\frac{7!}{2!1!3!1!} = \frac{7\,(6)\,(5)\,(4)\,(3)\,(2)\,(1)}{2\,(1)\,(3)\,(2)\,(1)\,(1)} = 420$$

Theorem3 The number of permutations of n distinct objects taken r at a time is

$${}_nP_r = \frac{n!}{(n-r)!}$$

Example The number of permutations of a, b, c taking 2 letters at a time is

$$\frac{3!}{(3-2)!} = 6$$

They are ab,ac, bc to which we add ba, ca, cb

Theorem4 The number of combinations of n distinct objects taken r at a time is

$$\binom{n}{r} = \frac{n!}{r!\,(n-r)!}$$

Example The number of combinations of a, b, c taking 2 letters at a time is

$$\frac{3!}{2!\,(3-2)!} = 3$$

They are ab,ac, bc

Example The number of possible hands of 5 cards each in a set of 52 different cards is

$$\begin{aligned}\binom{52}{5} &= \frac{n!}{r!\,(n-r)!} \\ &= \frac{52\,(51)\,(50)\,(49)\,(48)}{5\,(4)\,(3)\,(2)\,(1)} \\ &= 2,598,960\end{aligned}$$

1.5 Probability of an event

Probability is a function defined on the events in the sample space S in such a way that it obeys the following 3 axioms

1. $P(A) \geq 0$ for all sets A in S
2. $P(S) = 1$
3. If $A_1, A_2, ...,$ is a sequence of mutually exclusive events, then

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

Hence, if an experiment can result in N different equally likely outcomes and if A contains n outcomes, then the probability of A is

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

We make use of the counting principles to calculate the numerator and denominator.

Example A pair of dice is rolled. What is the probability of getting either a 7 or 11?

$$\begin{aligned} P("7" \cup "11") &= P("7") + P("11") \\ &= \frac{6}{36} + \frac{2}{36} = \frac{8}{36} \end{aligned}$$

Example Toss a fair coin 3 times. What is the probability of getting at least one head? The complementary event is getting no heads i.e. 3 tails.

$$\left(\frac{1}{8}\right)$$

Hence the probability of getting at least one head is $1 - \frac{1}{8} = \frac{7}{8}$

Example Poker hands consist of 5 cards drawn at random from a deck of 52 cards. The probability of getting a hand of 5 cards is

$$\binom{52}{5}^{-1}$$

What is the probability of getting four of a kind in the game of poker?

We can select one of the 13 possible types of cards (1,2,...,10,Jack, Queen, King) in 13 ways and then take all four cards. The fifth card can be chosen in 48 ways. Hence the required probability is

$$\frac{13\,(48)}{\binom{52}{5}} = \frac{1}{4165} = 2.4 \times 10^{-4}$$

Example What is the probability of a flush?

Here we want 5 cards of the same suit. We select the suit in $\binom{4}{1}$ ways and then choose 5cards from that suit in $\binom{13}{5}$ ways. Hence the probability of a flush is

$$\frac{\binom{4}{1}\binom{13}{5}}{\binom{52}{5}} = \frac{5148}{2598960} = 1.9808 \times 10^{-3}$$

1.6 Consequences of the axioms

From these axioms we can prove the following

- For any event A, $P(A') = 1 - P(A)$
- $P(\phi) = 0$
- $A_1 \subseteq A_2, \Longrightarrow P(A_1) \leq P(A_2)$
- If A and B are two events, then

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

- For three events A, B, C

$$\begin{aligned} P(A \cup B \cup C) &= P(A) + P(B) + P(C) \\ &- P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C) \end{aligned}$$

Examples We have the following probabilities for two events A_1, A_2: $P(A_1) = 0.12, P(A_2) = 0.30, P(A_1 \cap A_2) = 0.07$. Find $P(A_1 \cap A_2'), P(A_1' \cap A_2')$

From A Venn diagram, we see that $A_1 = (A_1 \cap A_2) \cup (A_1 \cap A_2')$

Hence $P(A_1) = P(A_1 \cap A_2) + P(A_1 \cap A_2') \Longrightarrow P(A_1 \cap A_2') = 0.12 - 0.07 = 0.05$

By De Morgan's laws,

$$\begin{aligned} P(A_1' \cap A_2') &= 1 - P(A_1 \cup A_2) \\ &= 1 - [P(A_1) + P(A_2) - P(A_1 \cap A_2)] \\ &= 1 - [0.12 + 0.30 - 0.07] \\ &= 0.65 \end{aligned}$$

Example A box contains 5 defective items and 20 non defectives. We sample 2 items at random without replacement. What is the probability that both are defective?

We have $(4)(5) = 20$ ways of choosing the 2 defectives in sampling without replacement. We have a total of $(25)(24) = 600$ ways of choosing 2 items without replacement. Hence the required probability is $\frac{20}{600} = 0.033$. Similarly the probability of having no defectives is

$$\frac{20(19)}{600} = 0.637$$

and the probability of having exactly one defective is $1 - 0.033 - 0.637 = 0.33$.

Example A man has 5 keys all similar in appearance. Only one opens his house door. He tries one key at a time selecting a key at random each time. What is the probability that the correct one is the third one chosen?

For the numerator, we count the number of ways of filling 5 boxes with a key. For box #3, there is only one way to choose the correct one. For the others, it is the product, $4(3)(2)(1)$. For the denominator it is the product $(5)(4)(3)(2)(1.)$ The ratio of the these products is $\frac{1}{5}$.

1.7 Conditional probability

There are many instances when the probability of an event may need to take into account the occurrence of another event. As an example, suppose that a balanced die is tossed once. The probability of getting the digit "2" is $1/6$. Suppose however we are told that the result of the toss was an even number. What then is the probability of getting the digit "2"?

If we know that the toss resulted in an even number, then the "new" sample space has possible outcomes $2, 4, 6$. Hence, given this information, the probability of getting the digit "2" is $1/3$.

Formally, we may obtain the conditional probability from the original probability function defined on the events of S.

Definition The conditional probability of B given A is

$$P(B|A) = \frac{P(A \cap B)}{P(A)}, P(A) > 0$$

Hence, $P(A \cap B) = P(A)\,P(B|A), P(A) > 0$

Application Consider the 2×2 table used to study cancer rates among men and women. A sample of 500 men and 500 women was taken.

	Cancer	**No cancer**	**Total**
Male	**60**	**440**	**500**
Female	**40**	**460**	**500**
Total	**100**	**900**	**1000**

If A is the event that an individual chosen at random from the group of 1000 is Male and B is the event that he has Cancer, then we see that

$$\begin{aligned}
P(A) &= \frac{500}{1000} = 0.50 \\
P(B) &= \frac{100}{1000} = 0.10 \\
P(A \cap B) &= \frac{60}{1000} = 0.06 \\
P(B|A) &= \frac{\left(\frac{60}{1000}\right)}{\left(\frac{500}{1000}\right)} = \frac{60}{500} = 0.12 \\
P(A|B) &= 0.60 \\
P(A'|B') &= 0.40 \\
P((A|B')) &= 0.49
\end{aligned}$$

The conditional probability obeys the axioms of probability as can be shown. Hence, we have the following properties:

- For any events A, B $P(A'|B) = 1 - P(A|B)$
- $P(\phi|B) = 0$
- $A_1 \subseteq A_2, \Longrightarrow P(A_1|B) \leq P(A_2|B)$
- If A,B, C are three events, then

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

Example Let A be the event that a sample of water from Lake Ontario contains mercury, B the event that it contains iron and C the event that it contains arsenic. Suppose we know that $P(A) = 0.32, P(B) = 0.16, P(C|A) = 0.45, P(C|A') = 0.02, P(A \cap B) = 0.28$. Find $P(B|A), P(C \cap A), P(C)$.

From the definition of conditional probability, $P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{0.28}{0.32} = 0.875$.

$P(C \cap A) = P(C|A)P(A) = 0.45(0.32)$,$P(C \cap A') = P(C|A')P(A') = 0.02(0.68)$= 0.015

$P(C) = P(C \cap A) + P(C \cap A') = 0.45(0.32) + 0.02(0.68) = 0.1576$

Example Three horses, labeled A,B,C are in a race and their chances of winning are $0.3, 0.5, 0.2$ respectively. If horse C is scratched, what is B's chance of winning?

We would like to calculate

$$
\begin{aligned}
P(B|A \cup B) &= \frac{P(B)}{P(A \cup B)} \\
&= \frac{0.5}{0.5 + 0.3} = \frac{5}{8}
\end{aligned}
$$

1.7.1 Independence

Two events A, B are said to be independent if and only if

$$P(B|A) = P(A)$$

Equivalently

$$P(A|B) = P(A)$$

It follows that A, B are independent if and only if

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

Example In the cancer example, the events A, B are dependent since

$$0.06 = P(A \cap B) \neq 0.50 (0.10) = P(A) P(B)$$

Example Consider an electrical system where two independent components are connected in series vs two independent components connected in parallel. For the series system, the probability that it works is given by

$$P(A_1 \cap A_2) = P(A_1)\,P(A_2)$$

whereas for the parallel system, the probability that it works is given by

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1)\,P(A_2)$$

Suppose now that $P(A_1) = P(A_2) = x$,and $0 < x < 1$ then

$$P(A_1 \cup A_2) = 2x - x^2$$

whereas

$$P(A_1 \cap A_2) = x^2$$

It follows that the parallel system is always more likely to function since

$$2x - x^2 > x^2.$$

1.7.2 Product rule

Theorem Let $A_1, ..., A_k$ be k events in S. Then

$$P(A_1 \cap ... \cap A_k) = P(A_1)\, P(A_2|A_1)\, P(A_3|A_1 \cap A_2)\, ...P(A_k|A_1 \cap A_2 \cap ... \cap A_{k-1})$$

Definition Let $A_1, ..., A_k$ be k events in S. They are said to be mutually independent provided

$$\begin{aligned} P(A_i \cap A_j) &= P(A_i)\, P(A_j)\,, i \neq j \\ P(A_i \cap A_j \cap A_k) &= P(A_i)\, P(A_j)\, P(A_k)\,, i \neq j \neq k \\ &... \\ P(A_1 \cap ... \cap A_k) &= P(A_1)\, P(A_2)\, P(A_3)\, ...P(A_k) \end{aligned}$$

1.8 Bayes' rule

Suppose that events $B_1, ..., B_k$ constitute a partition of the sample space S, that is, i) they are mutually disjoint and ii) their union is S. In the next theorem we see that we can express the probability of any event as a function of conditional probabilities on the events of the partition. In practice, this facilitates the calculation of the probability of the event.

Theorem For any other event A in S,

$$P(A) = \sum_{i=1}^{k} P(B_i \cap A) = \sum_{i=1}^{k} P(A|B_i) P(B_i)$$

Application Suppose we have an urn containing 3 red balls and 2 green balls. Consider the experiment whereby we pick a ball without replacement at random from the urn and note its color. We then select a second ball from the urn and note its color. What is the probability that the second ball drawn is red?

The answer depends on the color of the ball first chosen. Let B be the event that the first ball drawn is red. Then B and B' constitute a partition. If A represents the event that the second ball drawn is red, we have

$$\begin{aligned} P(A) &= P(A|B)\,P(B) + P(A|B')\,P(B') \\ &= \left(\frac{2}{4}\right)\left(\frac{3}{5}\right) + \left(\frac{3}{4}\right)\left(\frac{2}{5}\right) \\ &= \frac{12}{20} \end{aligned}$$

In Bayes' theorem, we go from the "effect" to the "cause".

Bayes'Theorem Suppose that events $B_1, ..., B_k$ constitute a partition of the sample space S such $P(B_i) > 0, i = 1, ..., k$. Then for any event A

$$\begin{aligned} P(B_r|A) &= \frac{P(B_r \cap A)}{\sum_{i=1}^{k} P(B_i \cap A)} \\ &= \frac{P(B_r)\, P(A|B_r)}{\sum_{i=1}^{k} P(B_i)\, P(A|B_i)} \end{aligned}$$

Application Consider the situation in a factory where there are 3 different machines producing a certain type of bolt. The machines have different levels of production and different probabilities of turning out defective bolts. At the end of the day, all the bolts produced are pooled together and a final inspection is made. Suppose that a bolt, selected at random, is found to be defective. What are the probabilities that it came from each of the 3 machines?

In this example, we see the "effect" which is to observe a defective bolt at the end of the day. We now wish to go back to the cause and trace which machine is most likely to have produced it.