MAT 2379 (Fall 2021)
(Introduction to Probability - Part I)

We can think of **statistics** as the study of data collection and analysis. Usually the data come from a random experiment or study.

Definition: A **random experiment** is an experiment whose outcome cannot be predicted with certainty.

Definition: The sample space, denoted S, for a random experiment is the set of all possible outcomes.

Example 1:
(a) Consider two groups of patients, smokers and non-smokers. Suppose that there are 12 smokers and 15 non-smokers. We decide to select one of the 27 patients at random. The latter experiment is a *random experiment* since we cannot predict with certainty the person that we will choose. A reasonable sample space for this experiment is

$$S = \{\text{patient } 1, \text{patient } 2, \ldots, \text{patient } 27\}.$$

(b) Diffusion of Molecules: Suppose that there are 10 molecules in a cell. After 10 minutes, we will count the number of molecules remaining in the cell among these 10 original molecules. The sample space is

$$S = \{0, 1, 2, \ldots, 10\}.$$

We might be interested in the likelihood of observing an outcome that satisfies a certain criterion or many criteria. For example, in the latter example, part (a), we may like to know the chances of selecting a smoker. We will use a notion of an **event** to distinguish some outcomes from others.

Definition: Consider a random experiment with the sample space S. A subset A of S, that is a sub-collection of the possible outcomes of the random experiment, is called an **event**.

Remarks 1:

- We will denote events with upper-case letters usually from the beginning (but not necessarily from the beginning) of the alphabet, e.g. $A, B, C, D, \ldots$
- If the outcome of the random experiment is x and x belongs to the event A, the we will say the A has **occurred**.
- The sample space S is called a **certain event.**
- The empty event $\emptyset$, i.e. no outcomes are found in $\emptyset$, is called an **impossible event.**

Example 2: Consider the random experiments from Example 1.
(a) Define the event $A =$"the patient is a smoker", that is

$$A = \{\text{smoker } 1, \text{smoker } 2, \ldots, \text{smoker } 12\}.$$

(b) Here are examples of events.

- A="less than 2 molecules", i.e. $A = \{0, 1\}$.
- B="more than 6 molecules", i.e. $B = \{7, 8, 9, 10\}$.
- C="at most 3 molecules", i.e. $C = \{0, 1, 2, 3\}$.
- D="at least 6", i.e. $D = \{6, 7, 8, 9, 10\}$.

Interpreting Probabilities

"It is a truth very certain that when it is not in our power to determine what is true we ought to follow what is most probable" - Descartes

Probability is the theory of measuring the likelihood or the chances that an event will occur.

We will present **3** different approaches that can be used to assign a probability to an event.

Subjective Approach: We use the interval $[0, 1]$ as a scale to assess the likelihood of the occurrence of an event. The closer we are to 1, the more confidence that we have that the event will occur. The closer we are to 0, the more confidence that we have that the event will not occur. The value given is given as a personal opinion.

Here are a few examples of the subjective approach:

(a) A doctor has diagnosed a disease. After considering a few factors the doctor must assess the probability that the disease can be cured. The doctor forms a personal opinion based on his experience and knowledge of the disease.

(b) What are the chances that you will submit all 4 assignments this semester? Based on your personal opinion you can assign a probability to this event.

Remarks:
(i) The advantage of the personal approach is that it can always be used.
(ii) The disadvantage of the personal approach is that it is subjective. We can ask different people for their personal opinions and obtain different answers.

Classical Approach: Concepts of probability have been around for thousands of years, but probability theory did not arise as a branch of mathematics until the mid-seventeenth century. The first objective definition of a probability is given below.

> **[Equally Likely Model]:** Consider a random experiment with sample space S and *equally likely* outcomes. The probability that the event A will occur is
>
> $$P(A) = \frac{n(A)}{n(S)},$$
>
> where $n(A)$ represents the total number of outcomes.

Remark: We will think of the **Random Selection** of an object from a collection of N distinct objects as being a random experiment with equally likely outcomes.

Exemple 3: We select a blood sample from a population of 20 samples of which 3 are contaminated. What is the probability that the sample will be contaminated?

Solution : Let A="Blood sample is contaminated", The total sample is the population of 20 samples. So $n(S) = 20$, and $n(A) = 3$, therefore
$P(A) = \frac{n(A)}{n(S)} = \frac{3}{20}$ The probability that the sample will be contaminated is $3/20$.

Problems with the model: The classical method requires that the experiment be broken down into equally likely outcomes.

- It is not always possible to do so.
- It is not always clear when outcomes should be considered as equally likely.
- It is possible to imagine of experiments with outcomes that should not be considered as equally likely. Here are examples:
    1. What is the probability that the next patient at the emergency room at a hospital is classified as critical? Is it reasonable to consider the categories of the triage as being equally likely? It is doubtful!!!
    2. Suppose we are studying the infestation of the mountain beetle pine in a certain forest in British Columbia. The beetles leave popcorn-shaped masses of resin, called "pitch tubes," on the trunk where beetle tunneling begins. Suppose that we choose a tree and count the number of pitch tubes. Is it reasonable to consider the possible outcomes of this experiment as being equally likely? It is doubtful!! Well there can be $0, 1, 2, 3, \ldots$ pitch tubes on a tree. It is doubtful that observing none is equally likely as observing 10, or 20, or 1000, or 100000, and so on.

Remark: Sometimes we do not necessarily have a finite number of possible results or it is not reasonable to assume that the results are equally likely. Furthermore, we might not know the sample space and so we are unable to construct a probability model with the equally likely model (sometimes called the classical approach). Here is an alternative approach.

Frequentist approach: Suppose that among the last 1000 patients at the emergency room, only 15 were classified as critical. We could construct an empirical model from these 1000 patients, by considering each of these patients as a trial of our experiment. We will interpret the relative frequency that a patient is classified as critical, here it is $15/1000 = 1.5\%$ as the approximate probability that the next patient will be classified as critical.

[**Frequentist approach**] Consider a random experiment with sample space S. Assume that we repeat the experiment n times. The probability that the event A will occur is

$$P(A) = \lim_{n\to\infty} \frac{f_n(A)}{n},$$

where $f_n(A)$ is the number of times (the frequency) that we observe the occurence of the event A among the n trials of the experiment.

Example 4:

(a) In the last 30 years there are been about 105 rattle snake bites in Ontario. Among these people only 2 have died. Suppose that someone is bitten by a rattle snake, what is the probability that this person will die? **Answer:** approximately $2/105 = 0.0190$ (Since $n = 105, f_n(A) = 2$ by $\frac{f_n(A)}{n} = \frac{2}{105}$).

(b) Of the last 500 patients that were admitted into the emergency room, only 10 were considered as extremely urgent. What is the probability that the next person, that is admitted into the emergency room, will be considered as extremely urgent. **Answer:** approximately $10/500 = 0.02$(Since $n = 500, f_n(A) = 10$ by $\frac{f_n(A)}{n} = \frac{10}{500}$).

Disadvantages of the Relative Frequency Model:

- To determine the probability of an event, we must repeat the experiment an infinite number of times or at least a very large number of times to obtain a good approximation.

Remarks:

- We have seen three approaches to interpret a probability. The most appropriate approach will depend on the context.
- We will eventually want to manipulate probabilities, and to do so, we will develop some probability rules. The difficulty with many approaches is that we must show that the rule can be used within each approach. It would be better to define a more general concept of probability that includes all of the approaches. The modern theory of probability is based on 3 axioms that we will discuss in Chapter 2. Any consequences of the axioms are called probability rules. These rules allow us to manipulate probabilities. They also show us how to consistently manipulate probabilities within a subjective approach.
- Before moving on to the probability axioms, we will see some examples of elementary genetics that is a field of applied probability.

Elementary Genetics (Section 1.2):

- Hereditary characteristics of an organism are determined by genes.
- An allele is one of a number of alternative forms of the same gene. Most organisms will two alleles of each gene.
- For exemple, the gene that determines the height of a pea takes on two forms:
  - T for tall
  - t for dwarfism
- The genotype constitutes the set of the alleles of an organism. For our pea example, the possible genotypes: TT, tt, Tt, Tt.
  N.B.: If the alleles are identical, e.g. TT or tt, then we say that the organism is **homozygous**. Otherwise, the organism is said to be **heterozygous**.
- Certain alleles are **dominante** and others are **recessive**: the dominant allele is expressed at the detriment of the recessive allele. The expression is called the **phenotype**. For the pea example, there are two phenotypes: TT, tT and Tt are expressed as **tall plants**, and tt is expressed as **dwarfism**.

A few laws of Elementary Genetics:

- Mendel's Law of Segregation states that every individual organism contains two alleles for each trait, and that these alleles segregate (separate) during meiosis such that each gamete contains only one of the alleles. An offspring thus receives a pair of alleles : one allele for each trait from each parent.

  We will suppose that the parent gives to its offspring a copy of its gene with equal probability. For example, if the genotype of the mother is Aa, then her offspring will receive a or A, each with a probability of 0.5.
- Mendel's Law of Independent Assortment states that alleles for separate traits are passed independently of one another i.e from parents to offspring.

  In our problems involving Elementary Genetics, we will suppose that genes are not linked. (We will see later that this is not always a reasonable assumption.)

Example 5: Consider the crossing of two brown eyed parents, where B is the dominant allele for brown eyes and b is the recessive allele for blue eyes. Suppose that only the mother is heterozygous with respect to colour of their eyes. Both parents are heterozygous for hair colour (Hh), where H=brown hair is dominant over h=blond hair. In terms of both traits, the mother's genotype is $BbHh$ and the father's genotype is $BBHh$.

(a) Use a tree diagram to list of the possible genotypes for their child (Answer: Brown eye and Brown hair; Brown eye and blond hair).

(b) Compute the probability that their child will have brown eyes and blond hair.

Example 6:[Genes with multiple alleles] Refer to example 1.5 in the textbook. The gene I determines the blood type of an organism. The gene has three alleles I^A, I^B and i.

genotype	phenotype (blood type)
I^AI^A or I^Ai	A
I^BI^B or I^Bi	B
I^AI^B	AB
ii	O

We cross a female with blood type A that is heterozygous with a male of blood type AB. AB.
(a) Use a Punnett square to draw a list of the possible genotypes for this crossing.
(b) Determine the probability that their child will be of blood type A.

female/male	$\frac{1}{2}I^A$	$\frac{1}{2}I^B$
$\frac{1}{2}I^A$	$\frac{1}{4}I^AI^A$(type A)	$\frac{1}{4}I^AI^B$(type AB)
$\frac{1}{2}i$	$\frac{1}{4}I^Ai$(type A)	$\frac{1}{4}I^Bi$(type B)

The possible genotypes for this crossing are A, AB, B, and the probability that their child will be of blood type A is 1/2.