Unit 2
Introduction to Probability

1. Why We Need Probability

In topic 1 (*Summarizing Data*), our scope of work was a limited one, *describe the data at hand*. We learned some ways to summarize.

- **Histograms, frequency tables, plots;**
- **Means, medians; and**
- **Variance, SD, SE, MAD, MADM.**

And previously, (*Course Introduction*), we acknowledged that, in our day to day lives, we have some intuition for probability.

- **What are the chances of winning the lottery? (*simple probability*)**
- **From an array of treatment possibilities, each associated with its own costs and prognosis, what is the optimal therapy? (*conditional probability*)**

<u>Chance</u> - In these, we were appealing to "chance". The notion of "chance" is described using concepts of probability and events. We recognize this in such familiar questions as:

- **What are the chances that a diseased person will obtain a test result that indicates the same? (*Sensitivity*)**
- **If the experimental treatment has no effect, how likely is the observed discrepancy between the average response of the controls and the average response of the treated? (*Clinical Trial*)**

To appreciate the need for probability, consider the conceptual flow of this course:

- **Initial lense - Describe the data at hand (*Topic 1 – Summarizing Data*)**

- **Enlarged lense – The data are a sample from a population (*Topic 3 – Populations and Samples*). Thus, we are relaxing our previous view of the data as a "given".**

- **Same enlarged lense – If this sample is one of a collection of equally likely samples that could have been obtained, then what is the likelihood of having obtained the particular sample that is ours? (*Topic 2 – Introduction to Probability*).**

- **Often (and for purposes of this course), the data at hand can be reasonably regarded as a simple random sample from a particular population distribution – eg. - Bernoulli, Binomial, Normal. (*Topic 4 – Bernoulli and Binomial Distribution, Topic 5 – Normal Distribution*).**

- **Estimation – We seek an understanding, through estimation, of the source population that gave rise to the observed data. Eg – we might want to estimate the value of the mean parameter (μ) of the population distribution that gave rise to our sample (*Topic 6 – Estimation and Topic 9-Correlation and Regression*).**

- **Hypothesis Testing - Also of interest are some tools for a formal comparison of competing explanations. (*Topic 7-Hypothesis Testing, Topic 8 –Chi Square Tests*)**

Now we have a better sense of why we need probability. Of interest are such things as:

- **If the population source is known, what are the chances of obtaining a particular outcome? A particular collection of outcomes? A particular sample of data?**

- **We will use the tools of probability in confidence interval construction.**

- **We will use the tools of probability in statistical hypothesis testing.**

2. Definition Probability Model

<u>Setting</u> -

- **The source population is assumed known.** **Eg - we might conceive of the source population as a big bucket of different colored balls that has been uniformly stirred.**

- **The sample is assumed to be a simple random sample; thus, it is just one of a collection of "equally likely' samples.** **Eg – imagine you have reached into the bucket with your hand and scooped out a handful of colored balls.**

- ***Note – An introduction to populations and samples is provided in Topic 3.***

<u>Question</u> -

If the available sample is representative of the source population, what are the "chances" of obtaining the observed values?

This is a <u>"frequentist"</u> approach to probability. It is not the only approach.

Alternative approaches -

- **<u>Bayesian</u> - "This is a fair coin. It lands "heads" with probability 1/2.**

- **<u>Frequentist</u> – "In 100 tosses, this coin landed heads 48 times".**

- **<u>Subjective</u> - "This is my lucky coin".**

Probabilities and probability distributions are nothing more than extensions of the ideas of relative frequency and histograms, respectively:

Ignoring certain mathematical details, a discrete probability distribution consists of two ingredients:

1. **The possible values a random value can assume, together with**

2. **The probabilities with which these values are assumed.**

Note: An elaboration of this intuition is required for the definition of a continuous probability distribution.

Example -

- **Suppose the universe of all university students is known to include men and women in the ratio 53:47.**

- **Consider the random variable, X = gender of an individual student**
  For convenience, we will say

X = 0	**when the student is "male"**
X = 1	**when the student if "female"**

- **We have what we need to define a discrete probability distribution:**

Ingredient 1 - **Possible value of X is represented as x**	**Ingredient 2 -** **Probability [X = x]**
0 = male **1 = female** ↑ **Be sure to check that this enumeration of all possible outcomes is "exhaustive".**	**0.53** **0.47** ↑ **Be sure to check that these probabilities add up to 100% or a total of 1.00.**

More formally, probability can be defined as

- **the chance of observing a particular outcome (discrete), or**
- **the likelihood of an event (continuous).**

- **The concept of probability assumes a <u>stochastic</u> or <u>random</u> process: i.e., the outcome is not predetermined – there is an element of chance.**

- **In discussing *discrete* probabilities, we assign a <u>numerical weight</u> or "chances" to each outcome. This "chance of" an event is its likelihood of occurrence.**

<u>Notation</u> -

The probability of outcome O_i is denoted $P(O_i)$

- **The probability of each outcome is between 0 and 1, inclusive:**

$$0 <= P(O_i) <= 1 \text{ for all } i$$

- **Conceptually, we can conceive of a population as a collection of "elementary" events or sample points. (Eg – the population might be the collection of colored balls in the bucket mentioned earlier in these notes) Here, "elementary" is the idea that such an event cannot be broken down further. The probabilities of all possible elementary outcomes sum to 1.**

$$\sum_{\text{all possible elementary outcomes } O_i} P(O_i) = 1 \qquad \textbf{(\textit{something happens})}$$

- **An event E might be one or several elementary outcomes, O. If an event E is certain, then it occurs with probability 1. This allows us to write**

$$P(E) = 1.$$

- **If an event E is impossible,**

$$P(E) = 0.$$

Some More Formal Language

1. **(discrete case) A *probability model* is the set of assumptions used to assign probabilities to each outcome in the sample space. (Eg – in the case of the bucket of colored balls, the assumption might be that the collection balls has been uniformly mixed so that each ball has the same chance of being picked when you scoop out one ball)**

   The *sample space* is the universe, or collection, of all possible outcomes.

   Eg – the collection of colored balls in the bucket.

2. **A *probability distribution* defines the relationship between the outcomes and their likelihood of occurrence.**

3. **To *define a probability distribution*, we make an assumption (the probability model) and use this to assign likelihoods.**

   Eg – Suppose we imagine that the bucket contains 50 balls, 30 green and 20 orange. Next, imagine that the bucket has been uniformly mixed. If the game of "sampling" is "reach in and grab ONE", then there is a 30/50 chance of selecting a green ball and a 20/50 chance of selecting an orange ball.

4. **When the outcomes are all equally likely, the model is called a *uniform probability model***

   In the pages that follow, we will be working with this model and then some (hopefully) straightforward extensions.

   From there, we'll move on to probability models for describing the likelihood of a sample of outcomes where the chances of each outcome are not necessarily the same (Bernoulli, Binomial, Normal, etc).

3. The "Equally Likely" Setting Introduction to Probability Calculations

An "equally likely" setting is the game of rolling a die –

There are 6 possible outcomes: {1, 2, 3, 4, 5, 6}. The probability of each is:

P(1) = 1/6
P(2) = 1/6
. . .
P(6) = 1/6
Sum = 1

Another "equally likely" setting is the tossing of a coin –

There are 2 possible outcomes in the set of all possible outcomes {H, T}. Here, "H" stands for "heads" and "T" stands for "tails".

Probability Distribution:

O_i	$P(O_i)$
H	.5
T	.5
Sum =1	

<u>Here is another "equally likely" setting</u> –

The set of all possible samples of digits of sample size n that can be taken, with replacement, from a population of size N. E.g., for N=3, n=2:

Sample Space:
S = { (1,1), (1,2), (1,3), (2,2), (2,1), (2,3), (3,1), (3,2), (3,3) }

Probability Model:
Assumption: equally likely outcomes, with $N^n = 3^2 = 9$ outcomes

Probability Distribution:

Outcome, O_i	$P(O_i)$
(1,1)	1/9
(1,2)	1/9
…	…
(3,3)	1/9
	Sum =1

Note – More on the ideas of "with replacement" and "without replacement" later.

<u>Another "equally likely" setting</u> –

Toss 2 coins
Set of all possible outcomes: S = {HH, HT, TH, TT}

Probability Distribution:

Outcome, O_i	$P(O_i)$
HH	¼ = .25
HT	¼ = .25
TH	¼ = .25
TT	¼ = .25
	Sum =1

Introduction to Composite Events "E" –

Recall that we can also define composite events of interest, E, and compute their probabilities. Such composite events E are each composed of a set of elementary outcomes from the sample space. For the reason of their being "elementary", the probability of each event is the simple addition of the probabilities of the qualifying elementary outcomes:

Event, E	Set of Qualifying Outcomes, O	$P(E_i)$
E_1: 2 heads	{HH}	¼ = .25
E_2: Just 1 head	{HT, TH}	¼ + ¼ = .50
E_3: 0 heads	{TT}	¼ = .25
E_4: Both the same	{HH, TT}	¼ + ¼ = .50
E_5: At least 1 head	{HH, HT, TH}	¼ + ¼ + ¼ = .75

Soon, we will learn when we can simply add probabilities and when we cannot; the answer has to do with the idea of "mutually exclusive" which is introduced on page 16.

Another Example - Calculation of a Composite Event in An "Equally Likely" Setting –

Recall the set of all possible samples of size n=2 that can be taken, with replacement, from a population of size N=3:

Sample Space S of all possible elementary outcomes O:
S = { (1,1), (1,2), (1,3), (2,2), (2,1), (2,3), (3,1), (3,2), (3,3) }

Probability Model:
Each outcome is equally likely and is observed with probability 1/9

Probability Distribution:

Outcome, O_i	$P(O_i)$
(1,1)	1/9
(1,2)	1/9
…	…
(3,3)	1/9
	Sum =1

Suppose we are interested in the event that subject "2" is in our sample. Thus,

E: subject 2 is in the sample.

S={ (1,1), (1,2), (1,3), (2,1), (2,2), (2,3), (3,1), (3,2), (3,3) }

By the "red underlines" (or just "underlines" if you don't see the red), we notice that the event of interest occurs in 5 of the 9 equally likely samples

Pr{ E } = Pr { (1,2), (2,1), (2,2), (2,3), (3,2) } = 1/9 + 1/9 + 1/9 + 1/9 + 1/9 = 5/9 = .56

We have what we need to define the associated probability distribution.

1. **Sample Space**

2. **Elementary Outcomes or Sample Points**

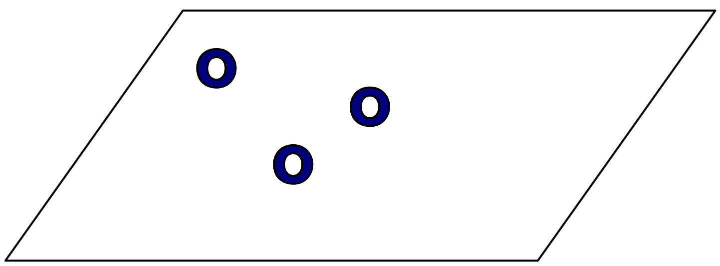
3. **Events**

4. Some Useful Basics

4a. Sample Space, Elementary Outcomes, Event

Population or Sample Space -

The population or sample space is defined as the set of all possible outcomes O of a random variable.



Example –

The sexes of the first and second of a pair of twins.
The population or sample space of all possible outcomes "O" is therefore:

{ boy, boy }
{ boy, girl }
{ girl, boy }
{ girl, girl },

The first sex is that of the first individual in the pair while the second sex is that of the second individual in the pair.

NOTE:
This random variable is a little different than most of the random variables described so far. Here, specification of one outcome requires two pieces of information (the sex of the first individual and the sex of the second individual) instead of one piece of information. It is an example of a bivariate random variable.

Elementary Outcome (Sample Point), O –

One sample point corresponds to each possible outcome "O" of a random variable. Such single points are elementary outcomes.

Example –

For the twins example, a single sample (elementary) point is:

{ boy, girl }

There are three other single sample (elementary) points: { girl, boy }, { girl, girl }, and { boy, boy }.

Event -

An event "E" is a collection of individual outcomes "O".

Notice that the individual outcomes or elementary sample points are denoted O_1, O_2, … while events are denoted E_1, E_2, …

Example –

Consider again the random variable defined as the sexes of the first and second individuals in a pair of twins. Consider, in particular, the event defined "boy". The set of outcomes defined by this event is the collection of all possible ways in which a pair of twins can include a boy. There are three such ways:

{ boy, boy }
{ boy, girl }
{ girl, boy }

Probability -

For now, we are considering a *discrete* probability distribution setting. Here, the probability of an event is the relative frequency with which ("chances of") at least one outcome of the event occurs in the population. If an event is denoted by E, then the probability that the event E occurs is written as P(E).

Example –

For the twins example, we might be interested in three events: E_1 = "two boys", E_2 = "two girls", E3 = "one boy, one girl". Assume that the chances of each sex are 1/2 for both the first and second twin and that the sex of the second twin is not determined in any way by the sex of the first twin. A table summarizing the individual probabilities of these events is:

Event	Description	Probability of Event
E_1	"two boys" {boy,boy}	Prob{1^{st}=boy}Prob{2^{nd}=boy} = (1/2) x (1/2) = 1/4 = 0.25
E_2	"two girls" {girl,girl}	Prob{1^{st}=girl}Prob{2^{nd}=girl} = (1/2) x (1/2) = 1/4 = 0.25
E_3	"one boy, one girl" {girl,boy} OR {boy,girl}	Prob{1^{st}=girl}Prob{2^{nd}=boy} + Prob{1^{st}=boy}Prob{2^{nd}=girl} = (1/2) x (1/2) = 1/4 = 0.25 + (1/2) x (1/2) = 1/4 = 0.25 = 0.50

4b. "Mutually Exclusive" and "Statistical Independence" Explained

The ideas of mutually exclusive and independence are different. The way we work with them are also different.

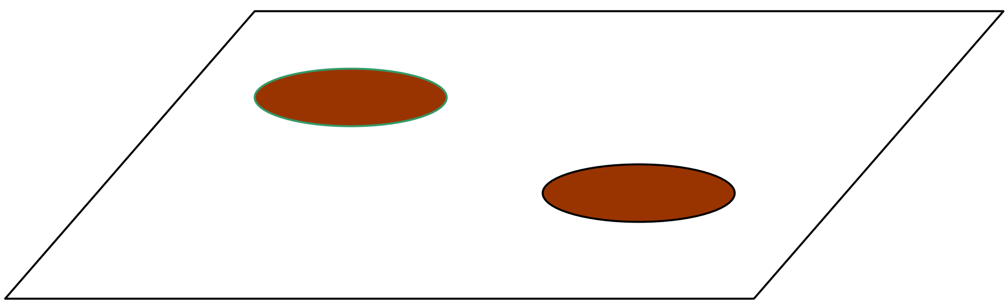
Mutually Exclusive ("cannot occur at the same time") -
Here's an example that illustrates the idea.

A coin toss cannot produce heads and tails simultaneously.

A coin landed "heads" excludes the possibility that the coin landed "tails".

We say the events of "heads" and "tails" in the outcome of a single coin toss are mutually exclusive.

Two events are mutually exclusive if they cannot occur at the same time.



Every day Examples of Mutually Exclusive –

- **Weight of an individual classified as "underweight", "normal", "overweight"**
- **Race/Ethnicity of an individual classified as "White Caucasian", "African American", "Latino", "Native African", "South East Asian", "Other"**
- **Gender of an individual classified as "Female", "Male", "Transgender"**

One More Example of Mutually Exclusive –

For the twins example introduced previously, the events E_1 = "two boys" and E_2 = "two girls" are mutually exclusive.

Statistical Independence -

To appreciate the idea of statistical independence, imagine you have in front of you two outcomes that you have observed. Eg.-

- **Individual is male with red hair.**
- **Individual is young in age with high blood pressure.**
- **First coin toss yields "heads" and second coin toss yields "tails".**

Two events are statistically independent if the chances, or likelihood, of one event is in no way related to the likelihood of the other event.

A familiar example of statistical independence corresponds to third example here - the first coin toss yielding heads and the second toss yielding tails. Under the assumption that the coin is "fair", we say

> **The outcomes on the first and second coin tosses are statistically independent and, therefore, probability ["heads" on 1st and "tails" on 2nd] = (1/2)(1/2) = 1/4.**

The distinction between "mutually exclusive" and "statistical independence" can be appreciated by incorporating an element of time.

> **Mutually Exclusive ("cannot occur at the same time")**
> **"heads on 1st coin toss" and "tails on 1st coin toss" are mutually exclusive.**
> **Probability ["heads on 1st" and "tails on 1st"] = 0**
>
> **Statistical Independence ("the first does not influence the occurrence of the later 2nd")**
> **"heads on 1st coin toss" and "tails on 2nd coin toss" are statistically independent**
> **Probability ["heads on 1st" and "tails on 2nd "] = (1/2) (1/2)**

SO … BE CAREFUL !!

The concepts of *mutually exclusive* and *independence* are distinct and in no sense equivalent. To see this:

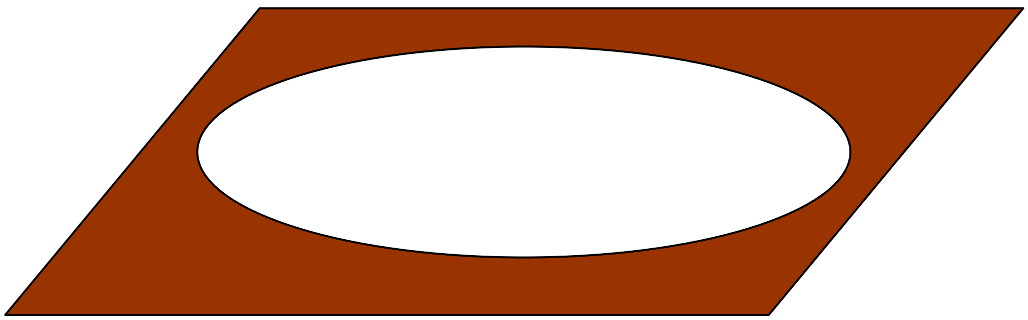
For A and B *independent*:
Pr(A and B) = P(A) P(B)
E.g. A=red hair B=winning the lottery

For A and B *mutually exclusive*:
Pr(A and B) = Pr (empty event) = 0
E.g. A=lottery winning is $10 B=lottery game not played

4c. Complement, Union, Intersection

Complement ("opposite", "everything else") -

The complement of an event E is the event consisting of all outcomes in the population or sample space that are *not* contained in the event E. The complement of the event E is denoted using a superscript c, as in E^c .



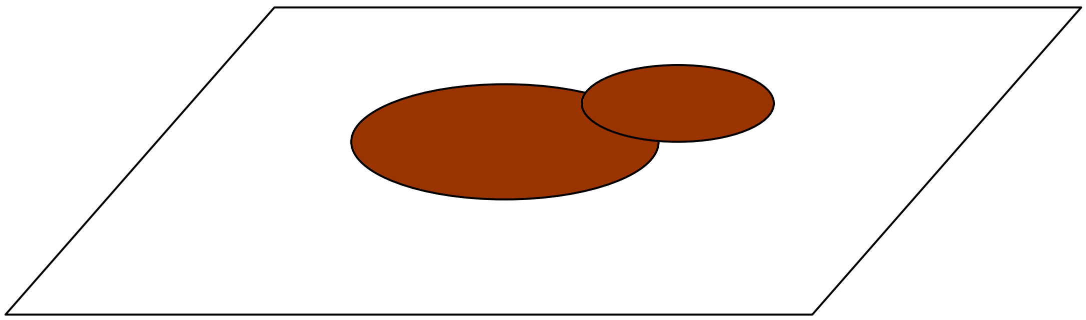
Example –

For the twins example, consider the event E_1 = "two boys". The complement of the event E_1 is

$$E_1^{\ c} = \{ \text{boy, girl} \}, \{ \text{girl, boy} \}, \{ \text{girl, girl} \}$$

Union, A or B ("either or") -

The union of two events, say A and B, is another event which contains those outcomes which are contained either in A or in B. The notation used is $A \cup B$.



Example –

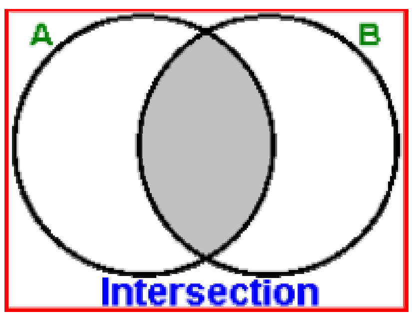
For the twins example suppose event "A" is defined { boy, girl } and event "B" is defined { girl, boy }. The union of events A and B is:

A $\cup$ B = { boy, girl }, { girl, boy }

Intersection, A and B ("only the common part") -

The intersection of two events, say A and B, is another event which contains only those outcomes which are contained in both A and in B. The notation used is A $\cap$ B.

Here it is depicted in the color gray.



Example –

For the twins example, consider next the events E_1 defined as "having a boy" and E_2 defined as "having a girl". Thus,

E_1 = { boy, boy }, { boy, girl }, { girl, boy }
E_2 = { girl, girl }, { boy, girl }, { girl, boy }

These two events do share some common outcomes.
The intersection of E_1 and E_2 is "having a boy and a girl":

$E_1 \cap E_2$ = { girl, boy }, { boy, girl }

5. Independence, Dependence, and Conditional Probability

Recall the illustration of statistical independence in the twins example –
The sex of the second twin was not determined in any way by the sex of the first twin. This is what is meant by independence.

The idea of dependence -
The occurrence of a first event alters, at least in part, the occurrence of a second event.

<u>Another Wording of Statistical Independence</u> -

Two events A and B are mutually independent if the chances, or likelihood, of one event is in no way related to the likelihood of the other event. When A and B are mutually independent:

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

Example –

Event A = "a woman is hypertensive"
Event B = "her mother-in-law is hypertensive".

The assumption of independence seems reasonable since the two women are not genetically related. If the probability of being hypertensive is 0.07 for each woman, then the probability that BOTH the woman and her mother-in-law are hypertensive is:

$$P(A \text{ and } B) = P(A) \times P(B) = 0.07 \times 0.07 = 0.0049$$

Dependence -

Two events are dependent if they are not independent.

The probability of both occurring depends on the outcome of at least one of the two.

Two events A and B are dependent if the probability of one event is related to the probability of the other:

$$P(A \text{ and } B) \neq P(A)\,P(B)$$

Example -

An offspring of a person with Huntington's Chorea has a 50% chance of contracting Huntington's Chorea. This is tantamount to saying *"If it is given that a person has Huntington's Chorea, then the chances of his/her offspring having the disease is 50%"*

Let Event A = "parent has Huntington's Chorea"
Let Event B = "offspring has Huntington's Chorea".

Without knowing anything about the parent's family background, suppose the chances that the parent has Huntington's Chorea is 0.0002.

Suppose further that, if the parent does not have Huntington's Chorea, the chances that the offspring has Huntington's Chorea is 0.000199.

However, if the parent has Huntington's Chorea, the chances that the offspring has Huntington's Chorea jumps to 0.50. Thus, the chances that both have Huntington's Chorea is not simply 0.0002 x 0.000199. That is,

$$P(A \text{ and } B) \neq P(A) \times P(B)$$

It is possible to calculate P(A and B), but this requires knowing how to relate P(A and B) to conditional probabilities. This is explained below.

Conditional Probability ("what happened first is assumed") -

Conditional probability refers to the probability of an event, given that another event is known to have occurred. We do these sorts of calculations all the time. As we'll see later in this course and in BE640, we might use the estimation of conditional probabilities when it is of interest to assess how dependent two events are, relative to each other.

The *conditional probability* that event B has occurred given that event A has occurred is denoted P(B|A) and is defined

$$P(B|A) = \frac{P(A \text{ and } B)}{P(A)}$$

provided that $P(A) \neq 0$.

Hint - ***When thinking about conditional probabilities, think in stages. Think of the two events A and B occurring chronologically, one after the other, either in time or space.***

Example - Huntington's Chorea again

- **The conditional probability that an offspring has Huntington's Chorea given a parent has Huntington's Chorea is 0.50.**
- **It is also known that the parent has Huntington's Chorea with probability 0.0002.**
- **Consider**

A = event that parent has Huntington's Chorea
B = event that offspring has Huntington's Chorea

- **Thus, we know**

Pr (A) = 0.0002
Pr (B|A) = 0.5

♣ **With these two "knowns", we can solve for the probability that both parent and child will have Huntington's Chorea. This is P(A and B)**

♣ $P(B|A) = \dfrac{P(A \text{ and } B)}{P(A)}$ **is the same as saying** $P(A \text{ and } B) = P(A)\,P(B|A)$

♣ **Probability both parent and child have Huntington's Chorea**

$$= P(A \text{ and } B)$$

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

$$= 0.0002 \times 0.5$$

$$= 0.0001$$

This type of staged probability calculation is often how probabilities of sequences of events are calculated. **An example is our understanding that cancer induction is the end result of several events (which must occur in the right sequence). Thus, the understanding is that probability of cancer is the product of Prob(event 1) x Prob (event 2 given event 1) x etc.**

6. Some Useful Tools for Calculating Probabilities

6a. The Addition Rule

A tool for calculating the probability of "either or both" of two events

Consider the following

Event A: **Go to Starbucks. Play the game A. A win produces a latte.
Suppose A yields win with probability = 0.15**

Event B: **Next go to Dunkin Donuts. Play the game B. A win produces a donut.
Suppose B yields "win" with probability = 0.35**

The Addition Rule

For two events, say A and B, the probability of an occurrence of either or both is written Pr [A ∪ B] and is

$$\Pr[A \cup B] = \Pr[A] + \Pr[B] - \Pr[A \text{ and } B]$$

Notice what happens to the addition rule if A and B are mutually exclusive!

$$\Pr[A \cup B] = \Pr[A] + \Pr[B]$$

Example –

In the scenario above, assume that the games offered by Starbucks and Dunkin Donuts are *statistically independent*. This means that

$$\Pr[A] = \Pr[\text{latte}] = 0.15$$
$$\Pr[B] = \Pr[\text{donut}] = 0.35$$
$$\Pr[A \text{ and } B] = \Pr[\text{latte \& donut}] = (0.15)(0.35) = 0.0525$$

We can use the addition rule to obtain the probability of winning a latte or a donut.

$$\Pr[\text{latte or donut}] = \Pr[A \cup B] = \Pr[A] + \Pr[B] - \Pr[A \text{ and } B]$$
$$= 0.15 + 0.35 - 0.0525$$
$$= 0.4475$$

6b. The Multiplication Rule

A tool for calculating the probability that "both" of two events occur

We consider again the Huntington's Chorea example and use the same scenario to illustrate the definition of the multiplication rule.

Event A: **A parent has probability of Huntington's Chorea = 0.0002**

Event B|A: **If it is known that a parent has Huntington's Chorea, then the conditional probability that the offspring also has Huntington's Chorea is = 0.50**

The Multiplication Rule

For two events, say A and B, the probability that both occur is written Pr [A ∩ B] and is

$$\Pr [A \cap B] = \Pr [A] \times \Pr [B \mid A]$$

Example –

In this scenario we have

Pr [A] = Pr [parent has disease] = 0.0002
Pr [B | A] = Pr [child has disease given parent does] = 0.50
Pr [A and B] = Pr [both have disease] = (0.0002)(0.50) = 0.0001

6c. Theorem of Total Probabilities

A tool for calculating the probability of a sequence of events

This is actually quite handy.

In thinking about a sequence of events, think in stages over time – that is, from one event to the next event and then to the next event after that, and so on. Now we'll play a game. The game has two steps.

Game -

Step 1: **Choose one of two games to play: G_1 or G_2**
G_1 is chosen with probability = 0.85
G_2 is chosen with probability = 0.15 (*notice that probabilities sum to 1)*

Step 2: **Given that you choose a game, G_1 or G_2 :**
G_1 yields "win" with conditional probability $P(win|G_1) = 0.01$
G_2 yields "win" with conditional probability $P(win|G_2) = 0.10$

What is the overall probability of a win, Pr(win)?

Hint – Think of all the distinct (mutually exclusive) ways in which a win could occur and sum their associated probabilities. There are only two such scenarios, and each scenario involves two component events.

$$Pr(win) = \underbrace{Pr[G_1 \text{ chosen}]}_{\text{event 1}} \ \underbrace{Pr[win|G_1]}_{\text{event 2}} + \underbrace{Pr[G_2 \text{ chosen}]}_{\text{event 1}} \ \underbrace{Pr[win|G_2]}_{\text{event 2}}$$

scenario #1 (first term) and **scenario #2** (second term)

$$= (.85)\ (.01) + (.15)\ (.10)$$

$$= 0.0235$$

This intuition has a name –
The Theorem of Total Probabilities.

Theorem of Total Probabilities

Suppose that a sample space S can be partitioned (carved up into bins) so that S is actually a union that looks like

$$S = G_1 \cup G_2 \cup \ldots \cup G_K$$

If you are interested in the overall probability that an event "E" has occurred, this is calculated

$$P[E] = P[G_1]P[E|G_1] + P[G_2]P[E|G_2] + \ldots + P[G_K]P[E|G_K]$$

provided the conditional probabilities are known.

Example – The lottery game just discussed.

G_1 = Game #1
G_2 = Game #2
E = Event of a win.

<u>So what? Applications of the Theorem of Total Probabilities</u> –

We'll see this again in this course and also in BE640

- ♣ **Diagnostic Testing**
- ♣ **Survival Analysis**

6d. Bayes Rule

A very useful combination of tools

Among our handy tools are the multiplication rule and the theorem of total probabilities.

1. **P(A and B) = P(A) P(B|A) = P(B) P(A|B)**

   This provides us with three ways of determining a joint probability

2. $P[E] = P[G_1]P[E|G_1] + P[G_2]P[E|G_2] + ... + P[G_K]P[E|G_K]$

   This provides us with a means of calculating an overall probability when things happen in an "ordered sequence" kind of way.

Putting "1" and "2" together provides us with another useful tool - Bayes Rule

Bayes Rule

Suppose that a sample space S can be partitioned (carved up into bins) so that S is actually a union that looks like

$$S = G_1 \cup G_2 \cup ... \cup G_K$$

If you are interested in calculating $P(G_i |E)$, this is calculated

$$P[G_i|E] = \frac{P(E|G_i)P(G_i)}{P[G_1]P[E|G_1] + P[G_2]P[E|G_2] + ... + P[G_K]P[E|G_K]}$$

provided the conditional probabilities are known.

Illustration of a Bayes Theorem Application

***Source:* http://yudkowsky.net/bayes/bayes.html You'll find a link to this URL on the course website page for *Topic 2: Introduction to Probability*. This URL is reader friendly.**

- **Suppose it is known that the probability of a positive mammogram is 80% for women with breast cancer and is 9.6% for women without breast cancer.**

- **Suppose it is also known that the likelihood of breast cancer is 1%**

- **If a women participating in screening is told she has a positive mammogram, what are the chances that she has breast cancer disease?**

Let

- **A = Event of breast cancer**
- **X = Event of positive mammogram**

What we want to calculate is Probability (A | X)

What we have as available information is

- **Probability (X | A) = .80** **Probability (A) = .01**
- **Probability (X | not A) = .096** **Probability (not A) = .99**

Here's how the solution works …

$$\Pr(\,A \mid X\,) = \frac{\Pr(A \text{ and } X)}{\Pr(X)}$$ **by definition of conditional Probability**

$$= \frac{\Pr(X \mid A)\Pr(A)}{\Pr(X)}$$ **because we can re-write the numerator this way**

$$= \frac{\Pr(X \mid A)\Pr(A)}{\Pr(X \mid A)\Pr(A) + \Pr(X \mid \text{not } A)\Pr(\text{not } A)}$$ **by thinking in steps in denominator**

$$= \frac{(.80)\,(.01)}{(.80)\,(.01) + (.096)\,(.99)}$$ **= .078 , representing a 7.8% likelihood.**

7. A Simple Probability Model: The Bernoulli Distribution

The Bernoulli Distribution is an example of a discrete probability distribution. It is an appropriate tool for the analysis of proportions and rates.

Recall the coin toss.

"50-50 chance of heads" can be re-cast as a random variable. Let

Z = random variable representing outcome of one toss, with

Z = 1 if "heads"
0 if "tails"

π= Probability [coin lands "heads" }. Thus,

$$\pi = \Pr[\ Z = 1\ ]$$

We have what we need to define a discrete probability distribution.

Ingredient 1 **Enumeration of all possible outcomes** - **outcomes are mutually exclusive** - **outcomes are exhaust all possibilities**	**1** **0**
Ingredient 2 **Associated probabilities of each** - **each probability is between 0 and 1** - **sum of probabilities totals 1**	**Outcome** — **Pr[outcome]** **0** — **$(1 - \pi)$** **1** — **π**

<u>In epidemiology, the Bernoulli distribution might be a model for the description of the outcome that is realized by ONE individual (N=1):</u>
This person is in one of two states. He or she is either in a state of:

1) **"event" with probability π** **("event" might be disease, mortality, etc)**
2) **"non event" with probability $(1-\pi)$**

The description of the likelihood of being either in the "event" state or the "non-event" state is given by the <u>Bernoulli distribution.</u> We'll use Z to represent this random variable outcome.

Bernoulli Distribution

Suppose Z can take on only two values, 1 or 0, and suppose:

Probability [Z = 1] = π
Probability [Z = 0] = $(1-\pi)$

This gives us the following expression for the likelihood of Z=z.

Probability [Z = z] = $\pi^z \ (1-\pi)^{1-z}$ for z=0 or 1.

In Topic 4, Bernoulli and Binomial Distribution, we'll discuss this distribution in more detail. See also Appendix 2 here.

<u>Example</u>: Z is the result of tossing a coin once. If it lands "heads" with probability = .5, then π = .5. Later we'll see that individual Bernoulli distributions are the basis of describing patterns of disease occurrence in a logistic regression analysis.

8. Probability in Diagnostic Testing

Students of epidemiology are introduced to, among other things:

- **concepts of diagnostic testing (sensitivity, specificity, predictive value positive, predictive value negative);**
- **concepts of disease occurrence (prevalence, incidence); and**
- **measures of association for describing exposure-disease relationships (risk, odds, relative risk, odds ratio).**

These have their origins in notions of conditional probability.

a. Prevalence ("existing")

The point prevalence of disease is the proportion of individuals in a population that has disease at a single point in time (point), regardless of the duration of time that the individual might have had the disease.

Prevalence is NOT a probability.

Example -

A study of sex and drug behaviors among gay men is being conducted in Boston, Massachusetts. At the time of enrollment, 30% of the study cohort were sero-positive for the HIV antibody. Rephrased, the point prevalence of HIV sero-positivity was 0.30 at baseline.

b. Cumulative Incidence ("new")

The cumulative incidence of disease is the probability an individual who did not previously have disease will develop the disease over a specified time period.

Example -

Consider again Example 1, the study of gay men and HIV sero-positivity. Suppose that, in the two years subsequent to enrollment, 8 of the 240 study subjects sero-converted. This represents a two year cumulative incidence of 8/240 or 3.33%.

c. Sensitivity, Specificity

The ideas of sensitivity, specificity, predictive value of a positive test, and predictive value of a negative test are most easily understood using data in the form of a 2x2 table:

		Disease Status: Present	Disease Status: Absent	
Test Result	**Positive**	a	b	a + b
	Negative	c	d	c + d
		a + c	b + d	a + b + c + d

In this table, a total of (a+b+c+d) individuals are cross-classified according to their values on two variables: disease (present or absent) and test result (positive or negative). It is assumed that a positive test result is suggestive of the presence of disease. The counts have the following meanings:

a = number of individuals who test positive AND have disease

b = number of individuals who test positive AND do NOT have disease

c = number of individuals who test negative AND have disease

d = number of individuals who test negative AND do NOT have disease

(a+b+c+d) = total number of individuals, regardless of test results or disease status

(b + d) = total number of individuals who do NOT have disease, regardless of their test outcomes

(a + c) = total number of individuals who DO have disease, regardless of their test outcomes

(a + b) = total number of individuals who have a POSITIVE test result, regardless of their disease status.

(c + d) = total number of individuals who have a NEGATIVE test result, regardless of their disease status.

Sensitivity

Among those persons who are known to have disease, what are the chances that the diagnostic test will yield a positive result?

To answer this question requires restricting attention to the subset of (a+c) persons who actually have disease. The number of persons in this subset is (a+c). Among this "restricted total" of (a+c), it is observed that "a" test positive.

$$\text{sensitivity} = \frac{a}{a+c}$$

Sensitivity is a conditional probability. It is the conditional probability that the test suggests disease given that the individual has the disease. For E_1=event that individual has disease and E_2=event that test suggests disease:

$$\textbf{sensitivity} = \mathbf{P(E_2 \mid E_1)}$$

To see that this is equal to what we think it should be, (a / [a+c]), use the definition of conditional probability:

$$P(E_2|E_1) = \frac{P(E_2 \text{ and } E_1)}{P(E_1)}$$

$$= \frac{[a/(a+b+c+d)]}{[(a+c)/(a+b+c+d)]}$$

$$= \left[\frac{a}{(a+c)}\right], \text{ which matches.}$$

Unfortunately, "sensitivity" also goes by other names:
- **positivity in disease**
- **true positive rate**

Specificity

Specificity pertains to:

Among those persons who do NOT have disease, what is the likelihood that the diagnostic test indicates this?

Specificity is a conditional probability. It is the conditional probability that the test suggests absence of disease given that the individual is without disease. For E_3=event that individual is disease free and E_4=event that test suggests absence of disease:

$$\text{specificity} = P(E_4 \mid E_3)$$

To see that this is equal to what we think it should be, (d / [b+d]), use the definition of conditional probability:

$$P(E_4|E_3) = \frac{P(E_4 \text{ and } E_3)}{P(E_3)}$$

$$= \frac{\left[d/(a+b+c+d)\right]}{\left[(b+d)/(a+b+c+d)\right]}$$

$$= \left[\frac{d}{(b+d)}\right]$$ **, which matches.**

"Specificity" also goes by other names:

* **negativity in health**
* **true negative rate**

d. Predictive Value Positive, Negative

Sensitivity and specificity are not very helpful in the clinical setting.

- **We don't know if the patient has disease (a requirement for sensitivity, specificity calculations).**
- **This is what we are wanting to learn.**
- **Thus, sensitivity and specificity are not the calculations performed in the clinical setting (they're calculated in the test development setting).**

Of interest to the clinician: "For the person who is found to test positive, what are the chances that he or she truly has disease?".

- **This is the idea of "predictive value positive test"**

"For the person who is known to test negative, what are the chances that he or she is truly disease free?".

- **This is the idea of "predictive value negative test"**

Predictive Value Positive Test

Among those persons who test positive for disease, what is the relative frequency of disease?

Predictive value positive test is also a conditional probability. It is the conditional probability that an individual with a test indicative of disease actually has disease. Attention is restricted to the subset of the (a+b) persons who test positive. Among this "restricted total" of (a+b),

$$\text{Predictive value positive} = \frac{a}{a+b}$$

Other Names for "Predictive Value Positive Test":

* **posttest probability of disease given a positive test**
* **posterior probability of disease given a positive test**

Also of interest to the clinician: Will unnecessary care be given to a person who does not have the disease?

Predictive Value Negative Test

Among those persons who test negative for disease, what is the relative frequency of absence of disease?

Predictive value negative test is also a conditional probability. It is the conditional probability that an individual with a test indicative of NO disease is actually disease free. Attention is restricted to the subset of the (c+d) persons who test negative. Among this "restricted total" of (c+d),

$$\text{Predictive value negative} = \frac{d}{c+d}$$

Other Names for "Predictive Value Negative Test":

* **posttest probability of NO disease given a negative test**
* **posterior probability of NO disease given a negative test**

9. Probability and Measures of Association for the 2x2 Table

Epidemiologists and public health researchers are often interested in exploring the relationship between a yes/no (dichotomous) exposure variable and a yes/no (dichotomous) disease outcome variable. A 2x2 summary table is again useful.

	Disease Status		
	Present	**Absent**	
Exposed	a	b	a + b
Not	c	d	c + d
	a + c	b + d	a + b + c + d

(a+b+c+d) = **total number of individuals, regardless of exposure or disease status**

(b + d) = **total number of individuals who do NOT have disease, regardless of their exposure status**

(a + c) = **total number of individuals who DO have disease, regardless of their exposure status**

(a + b) = **total number of individuals who have a POSITIVE exposure, regardless of their disease status.**

(c + d) = **total number of individuals who have a NO exposure, regardless of their disease status.**

a. Risk ("simple probability")

Risk of disease, without referring to any additional information, is simply the probability of disease. An estimate of the probability or risk of disease is provided by the relative frequency:

$$\frac{(a+c)}{(a+b+c+d)}$$

Typically, however, conditional risks are reported. For example, if it were of interest to estimate the risk of disease for persons with a positive exposure status, then attention would be restricted to the (a+b) persons positive on exposure. For these persons only, it seems reasonable to estimate the risk of disease by the relative frequency:

The straightforward calculation of the risk of disease for the persons known to have a positive exposure status is:

$$\text{P(Disease among Exposed)} = \frac{a}{(a+b)}$$

Repeating the calculation using the definition of conditional probability yields the same answer. Let E_1 =event of positive exposure and E_2=event of disease. Then:

Risk (disease given POSITIVE exposure) =

$$P(E_2|E_1) = \frac{P(E_2 \text{ and } E_1)}{P(E_1)}$$

$$= \frac{[a/(a+b+c+d)]}{[(a+b)/(a+b+c+d)]}$$

$$= \left[\frac{a}{(a+b)}\right]$$ **, which matches.**

b. Odds("comparison of two complementary (opposite) outcomes"):

In words, the odds of an event "E" is the chances of the event occurring in comparison to the chances of the same event NOT occurring.

$$\text{Odds} = \frac{\text{Pr(Event occurs)}}{\text{Pr(Event does NOT occur)}} = \frac{P(E)}{1 - P(E)} = \frac{P(E)}{(E^c)}$$

Example -

Perhaps the most familiar example of odds is reflected in the expression "the odds of a fair coin landing heads is 50-50". This is nothing more than:

$$\text{Odds(heads)} = \frac{P(\text{heads})}{P(\text{heads}^c)} = \frac{P(\text{heads})}{P(\text{tails})} = \frac{.50}{.50}$$

Similarly, for the exposure-disease data in the 2x2 table,

$$\text{Odds(disease)} = \frac{P(\text{disease})}{P(\text{disease}^c)} = \frac{P(\text{disease})}{P(\text{NO disease})} = \frac{(a+c)/(a+b+c+d)}{(b+d)/(a+b+c+d)} = \frac{(a+c)}{(b+d)}$$

$$\text{Odds(exposed)} = \frac{P(\text{exposed})}{P(\text{exposed}^c)} = \frac{P(\text{exposed})}{P(\text{NOT exposed})} = \frac{(a+b)/(a+b+c+d)}{(c+d)/(a+b+c+d)} = \frac{(a+b)}{(c+d)}$$

What if it is suspected that exposure has something to do with disease? In this setting, it might be more meaningful to report the odds of disease separately for persons who are exposed and persons who are not exposed. Now we're in the realm of <u>conditional odds</u>.

$$\text{Odds(disease | exposed)} = \frac{\text{Pr(disease|exposed)}}{\text{Pr(NO disease|exposed)}} = \frac{a/(a+b)}{b/(a+b)} = \frac{a}{b}$$

$$\text{Odds(disease | NOT exposed)} = \frac{\text{Pr(disease|not exposed)}}{\text{Pr(NO disease|not exposed)}} = \frac{c/(c+d)}{d/(c+d)} = \frac{c}{d}$$

Notice the vertical bar in expressions such as odds(disease|exposed). This vertical bar is nothing more than a "secretarial" shorthand that communicates a conditioning. Translation: "odds of disease given that exposure is present"

Similarly, one might calculate the odds of exposure separately for diseased persons and NON-diseased persons:

$$\text{Odds(exposed | disease)} = \frac{\text{Pr(exposed|disease)}}{\text{Pr(NOT exposed|disease)}} = \frac{a/(a+c)}{c/(a+c)} = \frac{a}{c}$$

$$\text{Odds(exposed | NO disease)} = \frac{\text{Pr(exposed|NO disease)}}{\text{Pr(NOT exposed|NO disease)}} = \frac{b/(b+d)}{d/(b+d)} = \frac{b}{d}$$

c. Relative Risk("comparison of two conditional probabilities")

Various epidemiological studies (prevalence, cohort, case-control designs) give rise to data in the form of counts in a 2x2 table.

Suppose we are interested in exploring the association between exposure and disease.

Recall our 2x2 table.

	Disease	Healthy	
Exposed	**a**	**b**	**a+b**
Not exposed	**c**	**d**	**c+d**
	a+c	**b+d**	**a+b+c+d**

Let's consider some actual counts, specifically 310 persons cross-classified by exposure and disease:

	Disease	Healthy	
Exposed	**2**	**8**	**10**
Not exposed	**10**	**290**	**300**
	12	**298**	**310**

We might have more than one 2x2 table if the population of interest is partitioned into subgroups or strata.

Example: Stratification by gender would yield a separate 2x2 table for men and women.

Relative Risk

The relative risk is the ratio of the conditional probability of disease among the exposed to the conditional probability of disease among the non-exposed.

Relative Risk: The ratio of two conditional probabilities

$$RR = \frac{a/(a+b)}{c/(c+d)}$$

Example: In our 2x2 table, we have a/(a+b) = 2/10 = .20, c/(c+d) = 10/300 = .0333
Thus, RR = .20 / .0333 = 6.006

- **It has been found empirically that many exposure-disease relationships vary with age in such a way that the log linear model is a good description. Specifically, the change with age in the relative risk of disease with exposure is reasonably stable. In such instances, the model is preferable to an additive risk model.**

d. Odds Ratio

The odds ratio measure of association has some wonderful advantages, both biological and analytical. Recall first the meaning of an "odds":

Recall that if p = Probability[event] then Odds[Event] = p/(1-p)

Let's look at the odds that are possible in our 2x2 table:

	Disease	Healthy	
Exposed	a	b	a+b
Not exposed	c	d	c+d
	a+c	b+d	a+b+c+d

Odds of disease among exposed $= \left[\dfrac{a/(a+b)}{b/(a+b)}\right] = \dfrac{a}{b} = \dfrac{2}{8} = .25$ **("cohort" study)**

Odds of disease among non exposed $= \left[\dfrac{c/(c+d)}{d/(c+d)}\right] = \dfrac{c}{d} = \dfrac{10}{290} = .0345$ **("cohort")**

Odds of exposure among diseased $= \left[\dfrac{a/(a+c)}{c/(a+c)}\right] = \dfrac{a}{c} = \dfrac{2}{10} = .20$ **("case-control")**

Odds of exposure among healthy $= \left[\dfrac{b/(b+d)}{d/(b+d)}\right] = \dfrac{b}{d} = \dfrac{8}{290} = .0276$ **("case-control")**

Students of epidemiology learn the following great result!

Odds ratio

In a cohort study:

$$OR = \frac{\text{Odds disease among exposed}}{\text{Odds disease among non-exposed}} = \frac{a/b}{c/d} = \frac{ad}{bc}$$

In a case-control study:

$$OR = \frac{\text{Odds exposure among diseased}}{\text{Odds exposure among healthy}} = \frac{a/c}{b/d} = \frac{ad}{bc}$$

Terrific!

The OR is the same, regardless of the study design, cohort (prospective) or case-control (retrospective)

Note: Come back to this later if this is too "epidemiological"!

Example: In our 2x2 table, a =2, b=8, c=10, and d=290 so the OR = 7.25. This is slightly larger than the value of the RR = 6.006.

<u>Thus, there are advantages of the Odds Ratio, OR</u>.

1. Many exposure disease relationships are described better using ratio measures of association rather than difference measures of association.

2. $OR_{\text{cohort study}} = OR_{\text{case-control study}}$

3. The OR is the appropriate measure of association in a case-control study.

- **Note that it is <u>not</u> possible to estimate an incidence of disease in a retrospective study. This is because we select our study persons based on their disease status.**

4. When the disease is rare, $OR_{\text{case-control}} \approx RR$

Appendix 1
Some Elementary Laws of Probability

A. Definitions:

1) **One sample point corresponds to each possible outcome of a random variable.**

2) **The sample space or population consists of all sample points.**

3) **A group of events is said to be exhaustive if their union is the entire sample space or population.**
Example - For the variable SEX, the events "male" and "female" exhaust all possibilities.

4) **Two events A and B are said to be mutually exclusive or disjoint if their intersection is the empty set.**
Example, one cannot be simultaneously "male" and "female".

5) **Two events A and B are said to be complementary if they are both mutually exclusive and exhaustive.**

6) **The events $E_1, E_2, ..., E_n$ are said to partition the sample space or population if:**

   (i) **E_i is contained in the sample space, and**
   (ii) **The event (E_i and E_j) = empty set for all $i \neq j$;**
   (iii) **The event (E_1 or E_2 or ... or E_n) is the entire sample space or population.**

In words: $E_1, E_2, ..., E_n$ are said to partition the sample space if they are pairwise mutually exclusive and together exhaustive.

7) If the events E_1, E_2, ..., E_n partition the sample space such that $P(E_1) = P(E_2) = ... P(E_n)$, then:

(i) $P(E_i) = 1/n$, for all i=1,...,n. This means that
(ii) the events E_1, E_2, ..., E_n are equally likely.

8) For any event E in the sample space: $0 <= P(E) <= 1$.

9) P(empty event) = 0. The empty event is also called the null set.

10) P(sample space) = P(population) = 1.

11) $P(E) + P(E^c) = 1$

B. Addition of Probabilities -

1) **If events A and B are mutually exclusive:**
   (i) **P(A or B) = P(A) + P(B)**
   (ii) **P(A and B) = 0**

2) **More generally:**
   P(A or B) = P(A) + P(B) - P(A and B)

3) **If events E_1, ..., E_n are all pairwise mutually exclusive:**
   P(E_1 or ... or E_n) = P(E_1) + ... + P(E_n)

C. Conditional Probabilities -

1) **P(B|A) = P(A and B) / P(A)**

2) **If A and B are independent:**
   P(B|A) = P(B)

3) **If A and B are mutually exclusive:**
   P(B|A) = 0

4) **P(B|A) + P(B^c|A) = 1**

5) **If P(B|A) = P(B|A^c):**
   Then the events A and B are independent

D. Theorem of Total Probabilities -

Let E_1, ..., E_k be mutually exclusive events that partition the sample space. The unconditional probability of the event A can then be written as a weighted average of the conditional probabilities of the event A given the E_i; i=1, ..., k:

$$P(A) = P(A|E_1)P(E_1) + P(A|E_2)P(E_2) + \ldots + P(A|E_k)P(E_k)$$

E. Bayes Rule -

If the sample space is partitioned into k disjoint events E_1, ..., E_k, then for any event A:

$$P(E_j|A) = \frac{P(A|E_j)P(E_j)}{P(A|E_1)P(E_1)+P(A|E_2)P(E_2)+\ldots+P(A|E_k)P(E_k)}$$

Appendix 2
Introduction to the Concept of Expected Value

We'll talk about the concept of "expected value" often; this is an introduction.

Suppose you stop at a convenience store on your way home and play the lottery. In your mind, you already have an idea of your chances of winning. That is, you have considered the question "what are the likely winnings?".

Here is an illustrative example. Suppose the back of your lottery ticket tells you the following–

\$1 is won with probability = 0.50
\$5 is won with probability = 0.25
\$10 is won with probability = 0.15
\$25 is won with probability = 0.10

THEN "likely winning" = [\$1](probability of a \$1 ticket) + [\$5](probability of a \$5 ticket) + [\$10](probability of a \$10 ticket) + [\$25](probability of a \$25 ticket)

=[\$1](0.50) + [\$5](0.25) +[\$10](0.15) + [\$25](0.10)

= \$5.75

Do you notice that the dollar amount \$5.75, even though it is called "most likely" is not actually a possible winning? What it represents then is a "long run average".

Other names for this intuition are

- ♣ **Expected winnings**
- ♣ **"Long range average"**
- ♣ **Statistical expectation!**

Statistical Expectation for a Discrete Random Variable is the Same Idea.

For a discrete random variable X (e.g. winning in lottery)
Having probability distribution as follows:

Value of X, x =	**P[X = x] =**
$ 1	**0.50**
$ 5	**0.25**
$10	**0.15**
$25	**0.10**

The random variable X has ***statistical expectation E[X]=μ***

$$\mu = \sum_{\text{all possible X=x}} [x]P(X = x)$$

Example –

In the "likely winnings" example, μ = $5.75

Statistical Expectation for the Bernoulli Distribution

Recall from this reading the introduction to the Bernoulli distribution. We can expand this introduction to include a discussion of its statistical expectation.

For Z distributed Bernoulli, we can calculate its statistical expectation. This statistical expectation has a name, the mean of the Bernoulli Distribution. It is represented as E[Z].

E[Z] = π because the following is true:

$$E[Z] = \sum_{\text{All possible z}} [z]\text{Probability}[Z = z]$$

$$= [0]\Pr[Z = 0] + [1] + \Pr[Z = 1]$$

$$= [0](1 - \pi) + [1](\pi)$$
$$= \pi$$

We can also calculate the expected value of $[Z - \pi]^2$. This statistical expectation also has a name, the variance of the Bernoulli Distribution .

Var[Z] = π(1-π) because the following is true:

$$\text{Var}[Z] = E[(Z - \pi)^2] = \sum_{\text{All possible z}} [(z - \pi)^2]\text{Probability}[Z = z]$$

$$= [(0 - \pi)^2]\Pr[Z = 0] + [(1 - \pi)^2] + \Pr[Z = 1]$$

$$= [\pi^2](1 - \pi) + [(1 - \pi)^2](\pi)$$

$$= \pi(1 - \pi)[\pi + (1 - \pi)]$$

$$= \pi(1 - \pi)$$

A useful convention to know

For a generic random variable X

- **The statistical expectation of X is E[X]**
  - **It is called the mean**
  - **It is also called the first moment**
  - **It is represented using the μ symbol**
  - **Intuitively, you can think of it as the "in the long run" average**
  - **For the Bernoulli distribution with parameter π, we saw mean = $\mu = \pi$**

- **Now consider the statistical expectation of $[X-\mu]^2$. This is $E([X-\mu]^2)$**
  - **It is called the variance**
  - **It is also called the second central moment**
  - **It is represented using the σ^2 symbol**
  - **For the Bernoulli, we saw variance = $\sigma^2 = \pi(1-\pi)$**