

Unit 2

Introduction to Probability

“Chance favours only those who know how to court her”

- Charles Nicolle

A weather report statement such as *“the probability of rain today is 50%”* is familiar and we have an intuition for what it means. But what does it really mean? Our intuitions about probability come from our experiences of relative frequencies of events. Counting! This unit is a very basic introduction to probability. It is limited to the **“frequentist” approach**. It’s important because we hope to generalize findings from data in a sample to the population from which the sample came.

Valid inference from a sample to a population is best when the probability framework that links the sample to the population is known.

Nature

Population/
Sample

Observation/
Data

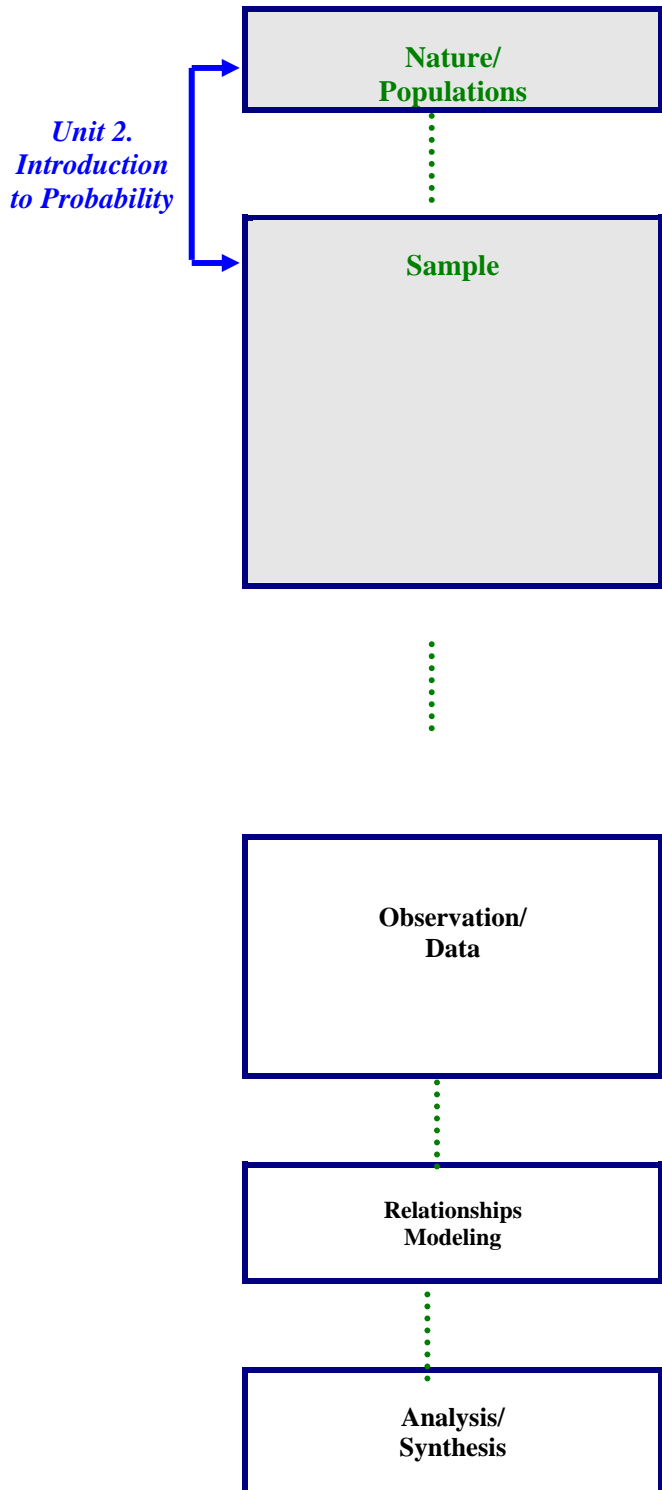
Relationships/
Modeling

Analysis/
Synthesis

Table of Contents

Topics	1. Unit Roadmap	3
	2. Learning Objectives	4
	3. Why We Need Probability	5
	4. <u>The Basics</u> : Some Terms Explained.....	7
	a. “Frequentist” versus “Bayesian” versus “Subjective”.....	7
	b. Sample Space, Elementary Outcomes, Events	8
	c. “Mutually Exclusive” and “Statistical Independence”	10
	d. Complement, Union, and Intersection	13
	e. The Joint Probability of Independent Events	15
	5. <u>The Basics</u> : Introduction to Probability Calculations	16
	“Equally Likely” Setting -	16
	6. The Addition Rule	23
	7. The Multiplication Rule - The Basics	26
	8. Conditional Probability	27
	a. Terms Explained: Independence, Dependence, Conditional Probability	27
	b. The Multiplication Rule – A little More Advanced	30
	c. TOOL - The Theorem of Total Probability	32
	d. TOOL - Bayes Rule	34
	9. <u>Probabilities in Practice</u> - Screening Tests and Bayes Rule	37
	a. Prevalence.....	37
	b. Incidence	37
	c. Sensitivity, Specificity	38
	d. Predictive Value Positive, Negative Test	41
	10. <u>Probabilities in Practice</u> - Risk, Odds, Relative Risk, Odds Ratio	43
	a. Risk	43
	b. Odds	45
	c. Relative Risk	47
	d. Odds Ratio	48
	Appendices	
	1. Some Elementary Laws of Probability	51
	2. Introduction to the Concept of Expected Value	54

1. Unit Roadmap



In Unit 1, we learned that a *variable* is something whose value can vary. A *random variable* is something that can take on different values *depending on chance*. In this context, we define a *sample space* as the collection of *all possible outcomes* for a random variable. **Note – It’s tempting to call this collection a ‘population.’ We don’t because we reserve that for describing nature. So we use the term “sample space” instead.**

An *event* is one outcome or a set of outcomes. Gerstman BB defines *probability* as the proportion of times an event is *expected* to occur in the population. It is a number between 0 and 1, with 0 meaning “never” and 1 meaning “always.”

Nature ——— Population/ Sample ——— Observation/ Data ——— Relationships/ Modeling ——— Analysis/ Synthesis

2. Learning Objectives

When you have finished this unit, you should be able to:

- Define probability, from a “frequentist” perspective.
- Explain the distinction between outcome and event.
- Calculate probabilities of outcomes and events using the proportional frequency approach.
- Define and explain the distinction among the following kinds of events: mutually exclusive, complement, union, intersection.
- Define and explain the distinction between independent and dependent events.
- Define prevalence, incidence, sensitivity, specificity, predictive value positive, and predictive value negative.
- Calculate predictive value positive using Bayes Rule.
- Defined and explain the distinction between risk and odds. Define and explain the distinction between relative risk and odds ratio.
- Calculate and interpret an estimate of relative risk from observed data in a 2x2 table.
- Calculate and interpret an estimate of odds ratio from observed data in a 2x2 table.

3. Why We Need Probability

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

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

And previously, (*Course Introduction*), we saw 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*)

Chance - In these settings, 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*)

Nature ——— Population/
Sample ——— Observation/
Data ——— Relationships/
Modeling ——— Analysis/
Synthesis

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

- **Initial lens** - We are given a set of data. We don't know where it came from. In summarizing it, our goal is to communicate its important features. (*Unit 1 – Summarizing Data*)
- **Enlarged lens** – Suppose now that we are interested in the population that gave rise to the sample. We ask how the sample was obtained. (*Unit 3 – Populations and Samples*).
- **Same enlarged lens** – If the sample was obtained using a method in which every possible sample is equally likely (this is called “simple random sampling”), we can then calculate the chances of particular outcomes or events (*Unit 2 – Introduction to Probability*).
- Sometimes, the data at hand can be reasonably regarded as a simple random sample from a particular population distribution – eg. - Bernoulli, Binomial, Normal. (*Unit 4 – Bernoulli and Binomial Distribution, Unit 5 – Normal Distribution*).
- **Estimation** – We use the values of the statistics we calculate from a sample of data as estimates of the unknown population parameter values from the population from which the sample was obtained. Eg – we might want to estimate the value of the population mean parameter (μ) using the value of the sample mean \bar{X} (*Unit 6 – Estimation and Unit 9-Correlation and Regression*).
- **Hypothesis Testing** – We use calculations of probabilities to assess the chances of obtaining data summaries such as ours (or more extreme) for a null hypothesis model relative to a competing model called the alternative hypothesis model. (*Unit 7-Hypothesis Testing, Unit 8 –Chi Square Tests*)

Going forward:

- We will use the tools of probability in [confidence interval construction](#).
- We will use the tools of probability in [statistical hypothesis testing](#).

Nature ——— Population/
Sample ——— Observation/
Data ——— Relationships/
Modeling ——— Analysis/
Synthesis

4. The Basics – Some Terms Explained

4a. “Frequentist” versus “Bayesian” versus “Subjective

A Frequentist Approach -

- Begin with a **sample space**: This is the collection of all possible outcome values of the random variable. **Example – A certain population of university students consists of 53 men and 47 women. Tip – Populations versus Sample Spaces:** A collection of all possible individuals in nature is termed a “population”. The collection of all possible outcomes of a random variable is termed a “sample space”
- Suppose that we can do sampling of the population by the method of simple random sampling many times. (more on this in Unit 3, Populations and Samples). Call the **event** of interest: “A” **Example, continued – Sampling is simple random sampling of 1 student. The event “A” is gender=female.**
- As the number of simple random samples drawn increases, the proportion of samples in which the event “A” occurs eventually settles down to a fixed proportion. This eventual fixed proportion is an example of a limiting value. **Example, continued – After 10 simple random samples, the proportion of samples in which gender=female might be 6/10, after 20 simple random samples, it might be 10/20, after 100 it might be 41/100, and so on. Eventually the proportion of samples in which gender=female will reach its limiting value of .47**
- The probability of event “A” is the value of the limiting proportion. **Example, continued – Probability [A] = .47**

Again - This is a “frequentist” approach to probability. It is just one of three approaches.

- **Bayesian** - “This is a fair coin. It lands “heads” with probability 1/2.
- **Frequentist** – “In 100 tosses, this coin landed heads 48 times”.
- **Subjective** - “This is my lucky coin”.




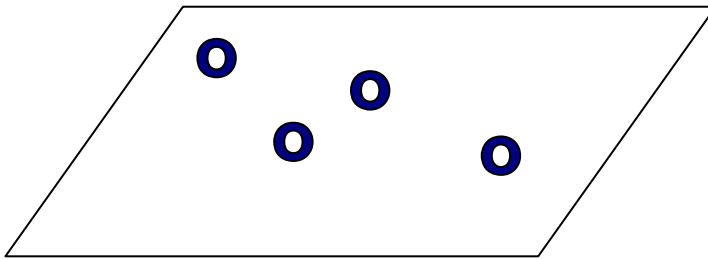
4b. Sample Space, Elementary Outcomes, Event

Sample Space -

The sample space is defined as the set of all possible (elementary) outcomes of a random variable.


I'm representing a sample space with a parallelogram that looks like this → 

I'm representing an (elementary) outcome with a dark circle that looks like this → 



Example –

The sexes of the first and second of a pair of twins.

The sample space of all possible outcomes contains 4 elementary outcomes  and they are:

- { 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.

Elementary Outcome (Sample Point), ○

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:

$$\text{○} = \{ \text{boy, girl} \}$$

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

Event -

Often, we are more interested in collections of elementary outcomes, as in the examples below. An event “E” is a collection of individual outcomes “O”.

Tip! Notice that the individual outcomes or elementary sample points are denoted O_1, O_2, \dots while events are denoted E_1, E_2, \dots

Example –

Consider again the random variable defined as the sexes of the first and second individuals in a pair of twins. Suppose we are interested in whether or not there is a boy. In the language of these notes, we now say that we are interested in the **event of “boy”**. The event of a boy is satisfied when any of three elementary outcomes has occurred. The three qualifying elementary outcomes are the following:

$$\begin{aligned} &\{ \text{boy, boy} \} \\ &\{ \text{boy, girl} \} \\ &\{ \text{girl, boy} \} \end{aligned}$$

Example – continued

Here are some more examples of “events”, together with their definitions:

<u>Notation</u>	<u>Event</u>	<u>Definition (“list of qualifying elementary outcomes”)</u>
E ₁	“boy”	{boy, boy} {boy, girl} {girl, boy}
E ₂	“two boys”	{boy, boy}
E ₃	“two girls”	{girl, girl}
E ₄	“one boy, one girl”	{boy, girl} {girl, boy}

Tip!

Like the familiar phrase “All salmon are fish, but not all fish are salmon”, we can also say “All elementary outcomes are events, but not all events are elementary outcomes”

4c. “Mutually Exclusive” and “Statistical Independence” Explained

Tip! Think time!

The ideas of **mutually exclusive** and **independence** are different. The ways we work with them are also different. The distinction can be appreciated by incorporating an **element of time**.

Mutually Exclusive (“cannot occur at the same time”) -

Here’s an example that illustrates the idea.

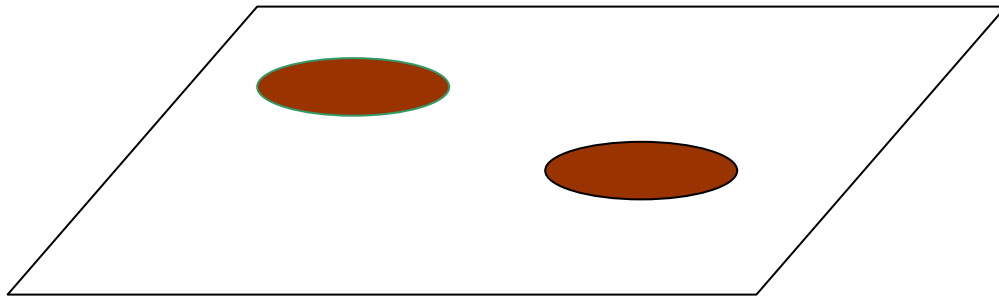
A **single** coin toss cannot produce heads and tails simultaneously.

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 $E_2 =$ “two boys” and $E_3 =$ “two girls” are mutually exclusive.

Statistical Independence -

Tip – Again, think time. Think about things occurring chronologically in time. 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.

We’ll get to calculating chances or likelihoods in the next section. But perhaps the following makes good sense to you anyway. Consider the game of tossing a single coin two times. The first coin toss lands heads and the second toss lands 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] = $(1/2)(1/2) = 1/4$.

Thus, 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

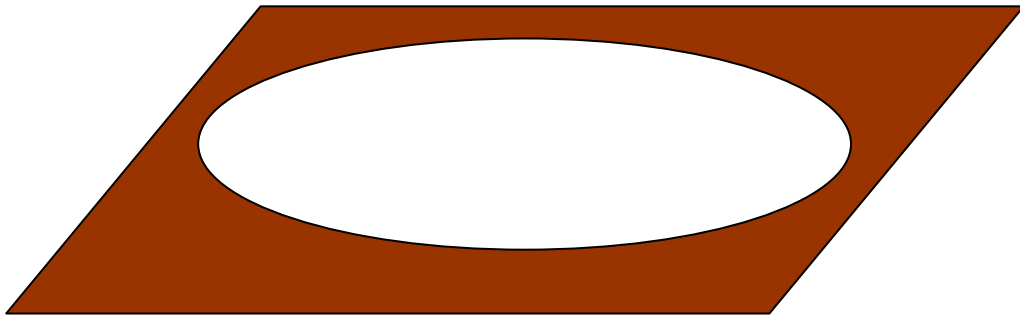
Probability [“heads on 1st” and “tails on 2nd”] = $(1/2) (1/2)$

Nature ——— Population/
Sample ——— Observation/
Data ——— Relationships/
Modeling ——— Analysis/
Synthesis

4d. 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 .



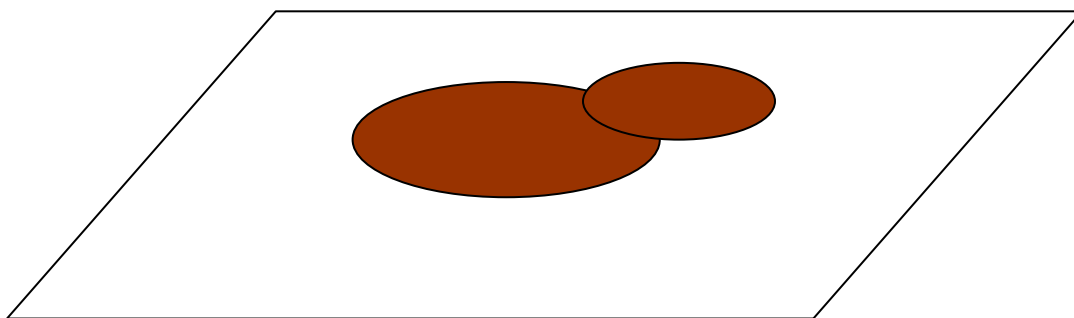
Example –

For the twins example, consider the event $E_2 =$ “two boys” that is introduced on page 10. The complement of the event E_2 is denoted E_2^c and is comprised of 3 elementary outcomes:

$$E_2^c = \{ \text{boy, girl} \}, \{ \text{girl, boy} \}, \{ \text{girl, girl} \}$$

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 –

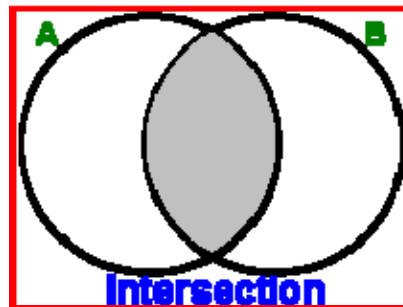
For the twins example, consider the event (this is actually an elementary outcome) {boy, girl}. Call this “A”. Call “B” the event {girl, boy}. The union of events A and B is:

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

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,

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

$$E_2 = \{ \text{girl, girl} \}, \{ \text{boy, girl} \}, \{ \text{girl, boy} \}$$

These two events do share some common outcomes. I’ve highlighted them in bold. The intersection of E_1 and E_2 is “having a boy and a girl”:

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

Nature

Population/
SampleObservation/
DataRelationships/
ModelingAnalysis/
Synthesis

4e. The Joint Probability of Independent Events

Independence, Dependence and conditional probability are discussed in more detail later. But we need a little bit of discussion of independent events here.

Specifically,

If the occurrence of event E_1 is independent of the occurrence of event E_2 ,

Then the probability that they BOTH occur is given by multiplying together the two probabilities:

$$\begin{aligned} \text{Probability [} E_1 \text{ occurs and } E_2 \text{ occurs]} \\ = \text{Pr}[E_1] \times \text{Pr}[E_2] \end{aligned}$$

Example –

The probability that two tosses of a fair coin both yield heads is $\frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$. chance.

E_1 = Outcome of heads on 1st toss. $\text{Pr}[E_1] = .50$

E_2 = Outcome of heads on 2nd toss $\text{Pr}[E_2] = .50$

$$\begin{aligned} \text{Pr[both land “heads”]} &= \text{Pr[heads on 1}^{\text{st}} \text{ toss AND heads on 2}^{\text{nd}} \text{ toss]} = \\ &= \text{Pr[heads on 1}^{\text{st}} \text{ toss]} \times \text{Pr[heads on 2}^{\text{nd}} \text{ toss]} \\ &= \text{Pr}[E_1] \times \text{Pr}[E_2] \\ &= (.50)(.50) \\ &= .25 \end{aligned}$$

5. The Basics – Introduction to Probability Calculations The Equally Likely Setting

In the “**frequentist**” framework, discrete probability calculations (these are also calculations of chance) are extensions of counting! We count the number of times an event of interest **actually occurred**, **relative to** how many times that event **could possibly have occurred**.

Example of an Equally Likely Setting -

- Consider again the **population** of 100 university students comprised of 53 men and 47 women.
- The experiment consists of selecting one university student (**elementary outcome**) **at random** from this population. “At random” here means that each of the 100 students has the same (1 in 100) chance of being selected. **Equally likely** refers to the common 1 in 100 chance of selection that each student has.
- We will define the outcome **event** of this experiment as the random variable, $X =$ gender of selected individual student. And, rather than having to write out “male” and “female”, we’ll use “0” and “1” as their representations, respectively. Thus, for convenience, we will say

$X = 0$ when the student is “male”
 $X = 1$ when the student is “female”

- We have what we need to define **probabilities** in an **equally likely setting**:

Sample Space of Events Random Variable X = Gender of Selected Individual	Probabilities of Event-Probability [X = x]
<div style="text-align: center;"> <p>0 = male 1 = female</p> <p>↑</p> </div> <p>Tip! Be sure to check that this enumeration of all possible outcomes is “exhaustive”.</p>	<div style="text-align: center;"> <p>0.53 0.47</p> <p>↑</p> </div> <p>Be sure to check that these probabilities add up to 100% or a total of 1.00.</p>

- **Tip - Just to be clear**
Equally likely refers to elementary outcomes – Each student has the same 1 in 100 chance of selection.



We’ve just developed a discrete probability calculation. More formally, discrete 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 stochastic or random process: i.e., the outcome is not predetermined – there is an element of chance.
- In discussing *discrete* probabilities, we assign a numerical weight or “chances” to each outcome. This “chance of” an event is its likelihood of occurrence.

Notation -

It’s tedious to write out long hand the probability of an event as Probability (O_i). So I may write this with a little shorthand. The following will all mean the same thing: (1) Probability (O_i), (2) Pr (O_i), and (3) $P(O_i)$

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

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

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

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

- Recall that an event E is defined as one or a set of elementary outcomes, O. If an event E is certain, then it occurs with probability 1. This allows us to write

$$P(E) = 1. \quad (\text{this event is always true ; eg – } E=\text{gravity exists})$$

- If an event E is impossible (meaning that it can never occur), we write

$$P(E) = 0. \quad (\text{this event is never true; eg – } E=\text{the earth is flat})$$



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 a 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 a 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).

Example 1 – Role ONE die exactly ONE time

There are 6 possible elementary outcomes: {1, 2, 3, 4, 5, 6}. They are all equally likely. Therefore, each has a 1 in 6 chance of occurrence:

$$\begin{aligned}
 P(1) &= 1/6 \\
 P(2) &= 1/6 \\
 &\dots \\
 P(6) &= 1/6 \\
 \text{Sum} &= 1 \text{ or } 100\%
 \end{aligned}$$

Example 2 – Toss ONE “fair” coin exactly ONE time

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 or 100%	

Example 3 – Select 1 digit from a list of size N. Return it. Then select a 2nd digit from the list of size N.

Tip – This is sampling with replacement

Suppose the list consists of the digits “1”, “2” and “3”.

- (1) Because there are 3 numbers in this list, $N=3$.
- (2) Because you sampled a digit 2 times, your sample size is $n=2$.
- (3) Because you did sampling with replacement, you had 3 choices for the first draw and 3 choices again for the second draw, there were $3 \times 3 = 3^2 = 9$ possible elementary outcomes.

$$3^2 = N^n = 9$$

$$N^n = [\text{length of list}] \text{ (number of times sampled)}$$



Sample Space

We'll denote the sample space of elementary outcomes using S:

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

Probability Model:

In this “equally likely” setting there are $N^n = 3^2 = 9$ outcomes, so each must have a 1 in 9 chance of occurrence:

Probability Distribution:

Outcome, O_i	$P(O_i)$
(1,1)	1/9
(1,2)	1/9
...	...
(3,3)	1/9
Sum =1 or 100%	

Tip – More on the ideas of “with replacement” and “without replacement” later.

Example 4 – Toss 2 coins

Sample Space

The sample space of all possible outcomes is $S = \{HH, HT, TH, TT\}$

Probability Distribution:

Outcome, O_i	$P(O_i)$
HH	$\frac{1}{4} = .25$
HT	$\frac{1}{4} = .25$
TH	$\frac{1}{4} = .25$
TT	$\frac{1}{4} = .25$
Sum =1 or 100%	



How to Calculate Probability [Events “E”] in the Equally Likely Setting

Answer – Add up the probabilities of the qualifying elementary outcomes.

Example 4, continued – Toss 2 coins

Event, E	Set of Qualifying Outcomes, O	P(Event, E _i)
E ₁ : 2 heads	{HH}	1/4 = .25
E ₂ : Just 1 head	{HT, TH}	1/4 + 1/4 = .50
E ₃ : 0 heads	{TT}	1/4 = .25
E ₄ : Both the same	{HH, TT}	1/4 + 1/4 = .50
E ₅ : At least 1 head	{HH, HT, TH}	1/4 + 1/4 + 1/4 = .75

Example 3, continued – Select 1 digit from a list of size N=3. Return it. Then select a 2nd digit from the list of size N. Tip – This is sampling with replacement

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 elementary 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 or 100%	



Event of Interest, E

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

E: subject 2 is in the sample.

Event, E	Set of Qualifying Outcomes, O	P(Event, E)
2 is in sample	{1,2} {2,1} {2,2} {2,3} {3,2}	$1/9 + 1/9 + 1/9 + 1/9 + 1/9 = .5555$

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



6. The Addition Rule

When to Use the Addition Rule

Tip – Think “or”

Use the addition rule when you want to calculate the event that **either “A” occurs or “B” occurs or both occur**.

Example 1 – Starbucks will give you a latte with probability = .15. Separately (meaning “independently”) Dunkin Donuts will give you a donut with probability = .35. What are the chances that you will get a latte OR a donut? Reminder - think “either or both” here

Event A: Go to Starbucks. Play game A. A win on game “A” produces a latte.
 So A yields win with probability = 0.15
 Thus, $Pr [A] = 0.15$

Event B: Next go to Dunkin Donuts. Play game B. A win on game “B” produces a donut.
 So B yields “win” with probability = 0.35
 Thus, $Pr [B] = 0.35$

Here is the formula with its solution. Justification follows ...

$$\begin{aligned}
 Pr[\text{latte OR donut}] &= Pr[\text{latte}] + Pr[\text{donut}] - Pr[\text{BOTH latte and donut}] \\
 &= 0.15 + 0.35 - (0.15)(0.35) \\
 &= 0.15 + 0.35 - 0.0525 \\
 &= 0.4475
 \end{aligned}$$

To see how this works, let’s first make a table of the elementary Outcomes and their Probabilities

Elementary Outcome, O	Probability of O, Pr[O]
{no latte, no donut}	$(.85)(.65) = .5525$ like two independent coin tosses, we multiply .85 x .65. More later!
{latte, no donut}	$(.15)(.65) = .0975$
{no latte, donut}	$(.85)(.35) = .2975$
{latte, donut}	$(.15)(.35) = .0525$
	Sum = 1 or 100%

We could do a brute force calculation of a “latte or donut” by summing the probabilities of the qualifying elementary outcomes. We get the correct answer = .4475

Event	Qualifying Elementary Outcomes	Probability of Qualifying Outcome
E = latte or donut or both	{latte, no donut}	.0975
	{no latte, donut}	.2975
	{latte, donut}	<u>.0525</u>
		Brute force sum = <u>.4475</u> matches ←

So why is there a subtraction in the formula

$$\Pr[\text{latte OR donut}] = \Pr[\text{latte}] + \Pr[\text{donut}] - \Pr[\text{BOTH latte and donut}]?$$

↑ That is - why is this subtraction done?

Answer – Because the elementary outcome {latte, donut} appears one time too many. It appears the event A=latte and the event B=donut. To avoid counting it twice, we subtract one of its appearances.

Event	Qualifying Elementary Outcome	Probability of Qualifying Outcome
A = latte	{latte, no donut}	.0975
	{latte, donut}	<u>.0525</u>
		Pr [A] = sum = .15

Event	Qualifying Elementary Outcomes	Probability of Qualifying Outcome
B= donut	{no latte, donut}	.2975
	+ {latte, donut}	<u>.0525</u>
		Pr[B] = sum = .35

Event	Qualifying Elementary Outcomes	Probability of Qualifying Outcome
Both	- {latte, donut}	.0525

$$\Pr[\text{latte OR donut}] = \Pr[\text{latte}] + \Pr[\text{donut}] - \Pr[\text{BOTH latte and donut}] = .15 + .35 - .0525 = .4475 matches ←$$

The Addition Rule

For two events, say A and B, the probability of an occurrence of either or both is written $\Pr [A \cup B]$ 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**! The subtraction of $\Pr [A \text{ and } B]$ disappears because mutually exclusive events can never happen which means their probability is zero ($\Pr [A \text{ and } B]=0$)

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

7. The Multiplication Rule – The Basics

When to Use the Multiplication Rule

Tip – Think “and”

Use the multiplication rule when you want to calculate the event that **event “A” occurs and then event “B” occurs and then event “C” occurs and so on.**

The multiplication rule was introduced previously. See page 15, (4d, [The Joint Probability of Independent Events](#)).

The Joint Probability of Two Independent Events is Obtained by Simple Multiplication

Consider again the example from page 23 – Latte’s and Donuts!! –

Now, instead of the event of “latte OR donut”, suppose we are interested in the event of “**latte AND donut**”. It’s just so much nicer to have both! We use the multiplication rule to obtain the probability of both (much like we did on page 15 when we considered the outcome of two coin tosses).

- A = Outcome of **latte** at Starbucks. $\Pr[A] = .15$
- B = Outcome of **donut** at Dunkin Donuts $\Pr[B] = .35$

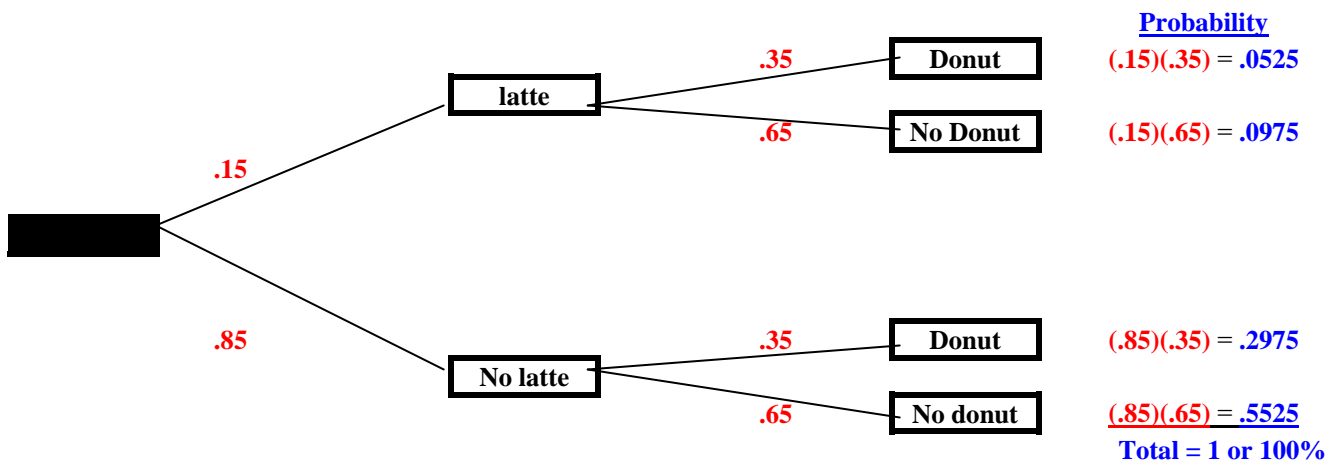
Under the assumption that the outcomes at Starbucks and Dunkin Donuts are independent of one another, we can obtain our answer by simple multiplication:

$$\Pr[\text{latte and donut}] = \Pr[\text{latte}] \times \Pr[\text{donut}] = (.15)(.35) = .0525$$

Introduction to Probability Trees

Latte’s and Donuts, continued –

We can construct a little probability tree that shows all the possibilities and their associated probabilities:



Thus, we see that the $\Pr[\text{latte and donut}] = .0525$ while $\Pr[\text{latte and no donut}] = .0975$, and so on.

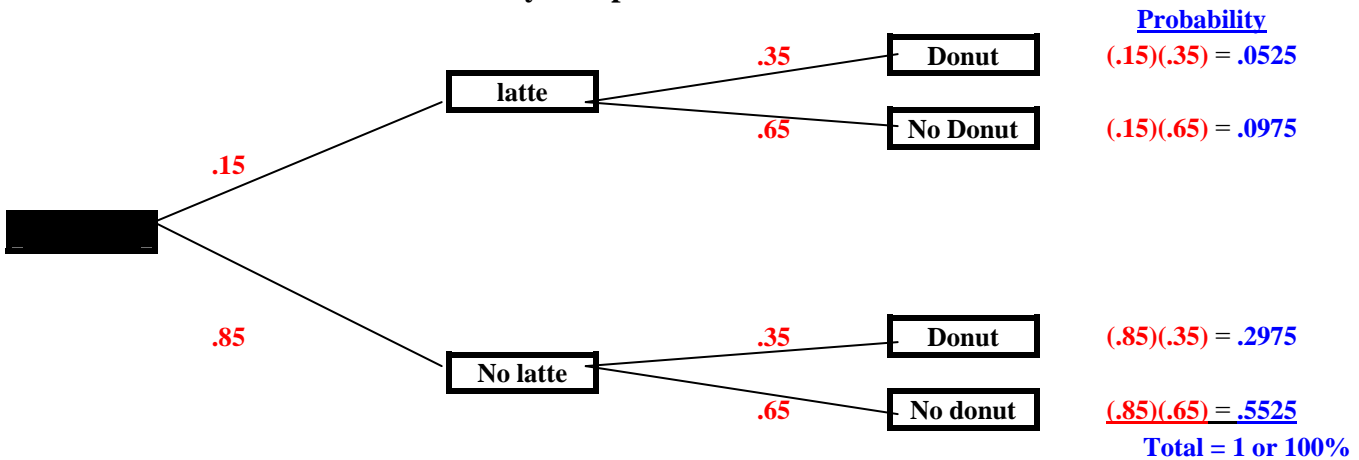
8. Conditional Probability

8a. Terms Explained: Independence, Dependence, and Conditional Probability

Statistical independence - The occurrence of Event A has **no influence** on the occurrence of Event B

In the “Latte, donut” example, the winning of a donut at Dunkin Donuts is not influenced in any way by the prior winning of a latte at Starbucks. You can see this in the probability tree. Regardless of a winning or not winning of a latte, the chances of a donut from Dunkin Donuts is the same .35 chance

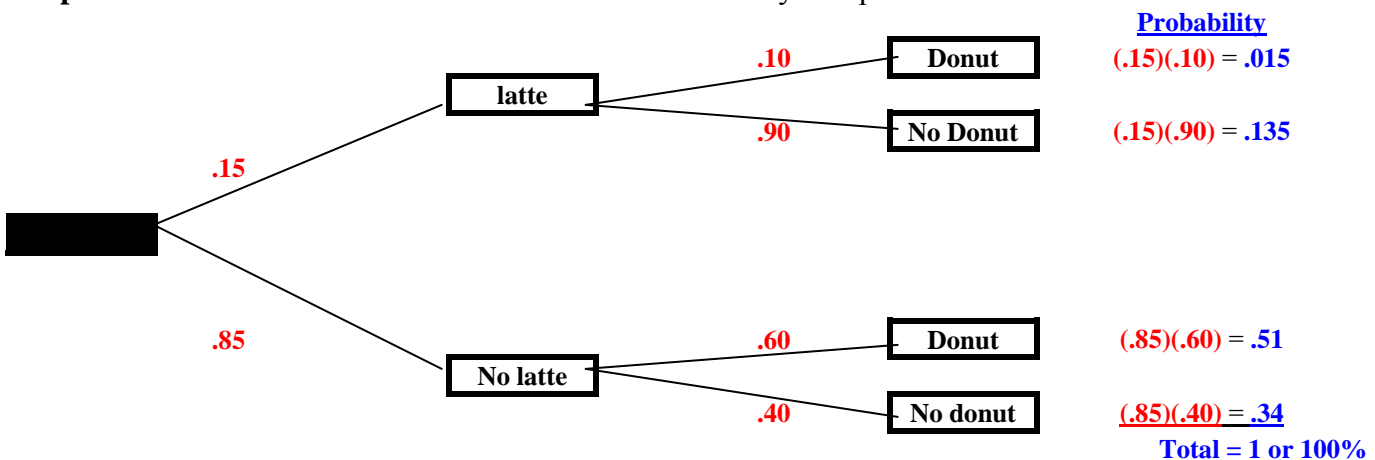
Latte and Donut Events are **Statistically Independent**



Dependence - The occurrence (or non-occurrence) of Event A **changes** the probability of occurrence of Event B

Now suppose that the Dunkin Donut policy on awarding donuts is different for the Starbucks latte winner than for the Starbucks latte loser. Suppose that Starbucks latte winners have only a 10% chance of winning a donut while Starbucks latte losers have a 60% chance of winning a donut. This is an example of dependence.

Dependence: Latte and Donut Events are **NOT** statistically Independent



Nature Population/
Sample Observation/
Data Relationships/
Modeling Analysis/
Synthesis

More on 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:

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

It is possible to calculate the joint probability of both “A” and “B occurring, $P(A \text{ and } B)$, but this requires knowing how to relate $P(A \text{ and } B)$ to what are called **conditional probabilities**. This is explained next.

Conditional Probability

Tip – You are in a situation where some prior event has just occurred. The particulars of this prior event influence the probability of the next event

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 $P(A) \neq 0$.

Example - Lattes and Donuts, again ...

- ♣ The conditional probability of winning a donut given you have previously won a latte is 0.10.
- ♣ The probability of having won the latte in the first place was 0.15.
- ♣ Consider

A = event of winning a latte at Starbucks

B = event of winning a donut at Dunkin Donuts

- ♣ Thus, we have

$$\Pr(A) = 0.15$$

$$\Pr(B|A) = 0.10$$

← This is a conditional probability namely, the conditional probability of a donut, conditional on (“given that”) a latte has been won

- ♣ With these two “knowns”, we can solve for the probability of winning both a latte and a donut. This is P(A and B)

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

- ♣ Probability of obtaining both latte and donut

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

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

$$= 0.15 \times 0.10$$

$$= 0.015$$

8b. The Multiplication Rule – A Little More Advanced

Tip – Think “and”

Now that we have the ideas of **probability trees** and **conditional probability**, we’re ready for the multiplication rule. Suppose you want to calculate the joint probability of **event “A” occurs and then event “B” occurs and then event “C” occurs and so on.**

Multiplication Rule

$$\Pr[A \text{ and } B \text{ and } C] = \Pr[A] \cdot \Pr[B|A] \cdot \Pr[C|B]$$

provided

- (1) $\Pr[A]$ is not zero; and
- (2) the conditional probabilities are known; and
- (3) the conditional probabilities are not zero.

Example – Survival analysis

Consider survival following an initial diagnosis of cancer. Suppose the one year survival rate is .75. Suppose further that, given survival to one year, the probability of surviving to two years is .85. Suppose then that given survival to two years, the probability of surviving five years is then .95. What is the **overall** two survival rate? What is the **overall** five year survival rate?

We can calculate the overall probabilities of surviving 2 and 5 years from just the probabilities given to us. The multiplication rule is used as follows:

Event of interest is probability of	What we are given are conditional probabilities
A = surviving 1 year	$\Pr [A] = .75$
B = surviving 2 years	$\Pr [B A] = .85$
C = surviving 5 years	$\Pr [C B] = .95$

Overall 2-year survival = $\Pr [B] = \Pr [A] \Pr [B | A]$

$\Pr [\text{survival is 2 years or more}]$
 = $\Pr[\text{surviving 1 year}] * \Pr[\text{surviving 2 years given survival to 1 year}]$
 = $\Pr[A] * \Pr[B | A]$
 = $(.75) * (.85)$
 = .6375



$$\begin{aligned} \text{Overall 5-year survival} &= \Pr [C] = \Pr [B] \Pr [C | B] \\ &= \Pr [A] \Pr [B | A] \Pr [C | B] \end{aligned}$$

$$\begin{aligned} \Pr [\text{survival is 5 years or more}] &= \Pr[\text{surviving 1 year}] * \Pr[\text{surviving 2 years given survival to 1 year}] * \Pr[\text{surviving 5 yrs given 2 yrs}] \\ &= \Pr[A] * \Pr[B | A] * \Pr[C | B] \\ &= (.75) * (.85) * (.95) \\ &= .605625 \end{aligned}$$

In PubHlth 640, you will learn this again as **Kaplan-Meier estimation**.

8c. TOOL - The Theorem of Total Probabilities

Tip! Think **chronology over time** and map it out using a **probability tree**.

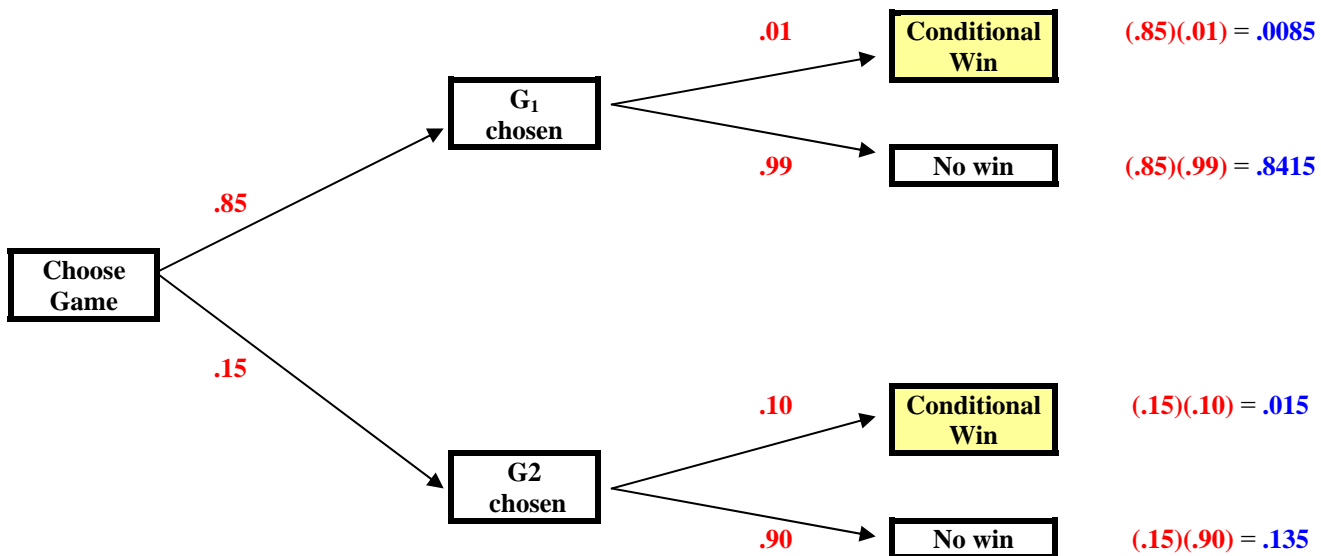
The theorem of total probabilities is used when there is **more than one path to a win** and you want to calculate the **overall probability of a win**.

Chronology over Time -

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 $P(\text{win}|G_1) = 0.01$
 G_2 yields “win” with conditional probability $P(\text{win}|G_2) = 0.10$

Probability Tree Tool -



What is the overall probability of a win, $\Pr(\text{win})$?

$$\begin{aligned} \Pr(\text{win}) &= \Pr[G_1 \text{ chosen}] \Pr[\text{win}|G_1] + \Pr[G_2 \text{ chosen}] \Pr[\text{win}|G_2] \\ &= (.85) (.01) + (.15) (.10) \\ &= .0085 + 0.015 \\ &= 0.0235 \end{aligned}$$

This is 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

$$S = G_1 \cup G_2 \cup \dots \cup G_k$$

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

$$P[E] = P[G_1]P[E|G_1] + P[G_2]P[E|G_2] + \dots + P[G_k]P[E|G_k]$$

provided the conditional probabilities are known.

Example – The lottery game just discussed.

E = Event of a win

G_1 = Game #1

G_2 = Game #2

We’ll see this again in this course and also in PubHlth 640

- ♣ Diagnostic Testing
- ♣ Survival Analysis

Nature

Population/
Sample

Observation/
Data

Relationships/
Modeling

Analysis/
Synthesis

8d. TOOL - Bayes Rule

Tip! Bayes Rule is a clever “putting together” of the **multiplication rule** and the **theorem of total probabilities**.

1. **Multiplication rule** – This provides us with a way to compute a joint probability

$$P(A \text{ and } B) = P(A) P(B|A) = P(B) P(A|B)$$
2. **Theorem of total probabilities** – This provides us with a way of computing an overall event

$$P[E] = P[G_1]P[E|G_1] + P[G_2]P[E|G_2] + \dots + P[G_k]P[E|G_k]$$

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

$$S = G_1 \cup G_2 \cup \dots \cup G_k$$

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

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

provided the conditional probabilities are known.

Example of Bayes Rule Calculation: “Given I have a positive screen, what are the chances that I have disease?”

Source: <http://yudkowsky.net/bayes/bayes.html> 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

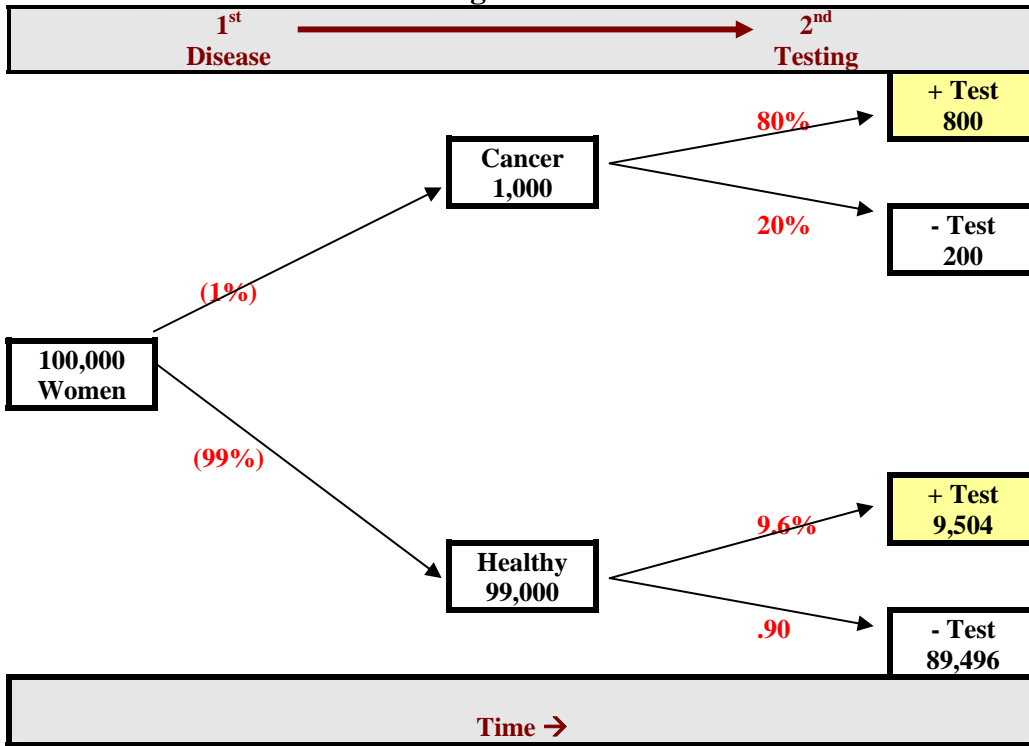
- Probability (X | A) = .80
- Probability (X | not A) = .096
- Probability (A) = .01
- Probability (not A) = .99

Here’s how the Bayes Rule solution works ...

$$\begin{aligned}
 \Pr(A | X) &= \frac{\Pr(A \text{ and } X)}{\Pr(X)} && \text{by definition of conditional Probability} \\
 &= \frac{\Pr(X | A) \Pr(A)}{\Pr(X)} && \text{because we can re-write the numerator this way} \\
 &= \frac{\Pr(X | A) \Pr(A)}{\Pr(X | A) \Pr(A) + \Pr(X | \text{not } A) \Pr(\text{not } A)} && \text{by thinking in steps in denominator} \\
 &= \frac{(.80) (.01)}{(.80) (.01) + (.096) (.99)} && = .078 , \text{ representing a 7.8\% likelihood.}
 \end{aligned}$$

Alternative to Bayes Rule Formula for the Formula Haters: “Given I have a positive screen, what are the chances that I have disease?”

Make a Tree of NUMBERS Using the Information about Prevalence and Test Results -



Counting Approach Solution

From the “tree” of numbers,

Total # with positive test = 800 + 9,504 = 10,304

of individuals with positive test and cancer = 800

Pr[cancer given positive test]

= Pr[cancer| + test]

= $\frac{\text{\# with cancer and +test}}{\text{\# with +test}}$

= $\frac{800}{10,304}$

= .078

which matches the answer on page 35!

9. Probabilities in Practice – Screening Tests and Bayes Rule

In your epidemiology class you are learning about:

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

Most of these have their origins in notions of **conditional probability**.

9a. 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; it is a proportion.

Example 1 -

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.

9b. Cumulative Incidence ("new")

The cumulative incidence is a probability. 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 2 -

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%.

Nature — Population/
Sample — Observation/
Data — Relationships/
Modeling — Analysis/
Synthesis

9c. 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		
		Breast Cancer	Healthy	
Test Result	Positive	800	9,504	10,304
	Negative	200	89,496	89,696
		1,000	99,000	100,000

		Disease Status		
		Present	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 (=800)
- b = number of individuals who test positive AND do NOT have disease (=9,504)
- c = number of individuals who test negative AND have disease (=200)
- d = number of individuals who test negative AND do NOT have disease (=89,496)

- (a+b+c+d) = total number of individuals, regardless of test results or disease status (= 100,000)
- (b + d) = total number of individuals who do NOT have disease, regardless of their test outcomes (= 99,000)
- (a + c) = total number of individuals who DO have disease, regardless of their test outcomes (= 1,000)
- (a + b) = total number of individuals who have a POSITIVE test result, regardless of their disease status. (= 10,304)
- (c + d) = total number of individuals who have a NEGATIVE test result, regardless of their disease status. (= 89,696)

Sensitivity



Sensitivity is a conditional probability. Among those persons who are known to have disease, what are the chances that the diagnostic test will **correctly** yield a positive result?

Counting Solution: From the 2x2 table, we would estimate sensitivity using the counts “a” and “c”:

$$\text{sensitivity} = \frac{a}{a + c} = \frac{800}{1,000} = .80$$

Conditional Probability Solution:

$$\begin{aligned} \text{sensitivity} &= \Pr[+\text{test}|\text{disease}] \\ &= \frac{\Pr[+\text{test AND "disease"}]}{\Pr[\text{disease}]} \\ &= \frac{[a/(a+b+c+d)]}{[(a+c)/(a+b+c+d)]} \\ &= \frac{[800/(100,000)]}{[(1,000)/(100,000)]} \\ &= \frac{800}{1,000} \\ &=.80 \end{aligned}$$

which matches the “counting” solution.

Specificity

Specificity is also conditional probability. Among those persons who are known to be disease free (“healthy”), what are the chances that the diagnostic test will **correctly** yield a negative result?

Counting Solution: From the 2x2 table, we would estimate sensitivity using the counts “b” and “d”:

$$\text{specificity} = \frac{d}{b+d} = \frac{89,496}{99,000} = .904$$

Conditional Probability Solution:

$$\text{specificity} = \Pr[-\text{test} | \text{NO disease}]$$

$$= \frac{\Pr["-\text{test}" \text{ AND } "NO \text{ disease}"]}{\Pr[\text{NO disease}]}$$

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

$$= \frac{[89,496/(100,000)]}{[(99,000)/(100,000)]}$$

$$= \frac{89,496}{99,000}$$

$$=.904$$

which matches the “counting” solution.

9d. 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

Predictive value positive test is a conditional probability. Among those persons who test positive for disease, what is the relative frequency of disease?

$$\text{Predictive value positive} = \frac{a}{a + b} = \frac{800}{10,304} = .078$$

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

Predictive value positive test is a conditional probability. Among those persons who test negative for disease, what is the relative frequency of disease-free?

$$\text{Predictive value negative} = \frac{d}{c+d} = \frac{89,496}{89,696} = .99$$

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

10. Probabilities in Practice – Risk, Odds, Relative Risk, Odds Ratio

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. Our 2x2 summary table is again useful.

NOTE! Note that the rows now represent EXPOSURE status. We can do this because these numbers are hypothetical anyway. They are just for illustration.

		Disease Status		
		Breast Cancer	Healthy	
Exposure Status	Exposed	800	9,504	10,304
	Not exposed	200	89,496	89,696
		1,000	99,000	100,000

		Disease Status		
		Present	Absent	
Test Result	Exposed	a	b	a + b
	Not exposed	c	d	c + d
		a + c	b + d	a + b + c + d

- a = number of exposed individuals who have disease (=800)
- b = number of exposed individuals who do NOT have disease (=9,504)
- c = number of NON exposed individuals who have disease (=200)
- d = number of NON exposed individuals who do NOT have disease (=89,496)

- (a+b+c+d) = total number of individuals, regardless of exposure or disease status (= 100,000)
- (b + d) = total number of individuals who do NOT have disease, regardless of exposure (= 99,000)
- (a + c) = total number of individuals who DO have disease, regardless of exposure (= 1,000)
- (a + b) = total number of individuals who are EXPOSED regardless of their disease status (= 10,304)
- (c + d) = total number of individuals who are NOT exposed regardless of their disease status. (= 89,696)

10a. 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:

$$\text{Overall risk} = \frac{(a+c)}{(a+b+c+d)} = \frac{1,000}{100,000} = .01$$

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:

$$P(\text{Disease} \mid \text{Exposed}) = \frac{a}{(a+b)} = \frac{800}{10,304} = .078$$

Repeating the calculation using the definition of conditional probability yields the same answer.

$$\begin{aligned} & \Pr[\text{disease} \mid + \text{exposure}] \\ &= \frac{\Pr["\text{disease}" \text{ AND } "+ \text{exposure}"]}{\Pr[+ \text{exposure}]} \\ &= \frac{\left[\frac{a}{(a+b+c+d)} \right]}{\left[\frac{(a+b)}{(a+b+c+d)} \right]} \\ &= \frac{\left[\frac{800}{(100,000)} \right]}{\left[\frac{(10,304)}{(100,000)} \right]} \\ &= \frac{800}{10,304} \\ &= .078 \end{aligned}$$

, which matches.

10b. 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.

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

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:

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

For the exposure-disease data in the 2x2 table,

$$\begin{aligned} \text{Odds(disease)} &= \frac{P(\text{disease})}{P(\text{disease}^c)} = \frac{P(\text{disease})}{P(\text{NO disease})} = \frac{(a + c)/(a + b + c + d)}{(b + d)/(a + b + c + d)} = \frac{(a + c)}{(b + d)} \\ &= \frac{1,000}{99,999} = .00101 \end{aligned}$$

$$\begin{aligned} \text{Odds(exposed)} &= \frac{P(\text{exposed})}{P(\text{exposed}^c)} = \frac{P(\text{exposed})}{P(\text{NOT exposed})} = \frac{(a + b)/(a + b + c + d)}{(c + d)/(a + b + c + d)} = \frac{(a + b)}{(c + d)} \\ &= \frac{10,304}{89,696} = .1149 \end{aligned}$$

Conditional Odds.

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 conditional odds.

$$\begin{aligned}\text{Odds(disease | exposed)} &= \frac{\text{Pr(disease|exposed)}}{\text{Pr(NO disease|exposed)}} = \frac{a / (a + b)}{b / (a + b)} = \frac{a}{b} \\ &= \frac{800}{9,504} = .0084\end{aligned}$$

$$\begin{aligned}\text{Odds(disease | NOT exposed)} &= \frac{\text{Pr(disease|not exposed)}}{\text{Pr(NO disease|not exposed)}} = \frac{c / (c + d)}{d / (c + d)} = \frac{c}{d} \\ &= \frac{200}{89,496} = .0002\end{aligned}$$

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

$$\begin{aligned}\text{Odds(exposed | disease)} &= \frac{\text{Pr(exposed|disease)}}{\text{Pr(NOT exposed|disease)}} = \frac{a / (a + c)}{c / (a + c)} = \frac{a}{c} \\ &= \frac{800}{200} = 4\end{aligned}$$

$$\begin{aligned}\text{Odds(exposed | NO disease)} &= \frac{\text{Pr(exposed|NO disease)}}{\text{Pr(NOT exposed|NO disease)}} = \frac{b / (b + d)}{d / (b + d)} = \frac{b}{d} \\ &= \frac{9,504}{89,496} = .1062\end{aligned}$$

11c. 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.

New!

Example –

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

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

Suppose now that we have (a+b+c+d) = 310 persons cross-classified by exposure and disease:

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

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

$$RR = \frac{a/(a+b)}{c/(c+d)} = \frac{2/10}{10/300} = 6$$

10d. 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 $p = \text{Probability}[\text{event}]$ then $\text{Odds}[\text{Event}] = p/(1-p)$

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

$$\text{Odds of disease among exposed} = \left[\frac{a/(a+b)}{b/(a+b)} \right] = \frac{a}{b} = \frac{2}{8} = .25 \quad (\text{“cohort” study})$$

$$\text{Odds of disease among non exposed} = \left[\frac{c/(c+d)}{d/(c+d)} \right] = \frac{c}{d} = \frac{10}{290} = .0345 \quad (\text{“cohort”})$$

$$\text{Odds of exposure among diseased} = \left[\frac{a/(a+c)}{c/(a+c)} \right] = \frac{a}{c} = \frac{2}{10} = .20 \quad (\text{“case-control”})$$

$$\text{Odds of exposure among healthy} = \left[\frac{b/(b+d)}{d/(b+d)} \right] = \frac{b}{d} = \frac{8}{290} = .0276 \quad (\text{“case-control”})$$

Students of epidemiology learn the following great result!

Odds ratio, OR

In a cohort study:

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

In a case-control study:

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

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, we have

$$\text{OR} = \frac{a \cdot d}{b \cdot c} = \frac{(2)(290)}{(8)(10)} = 7.25$$

Thus, there are advantages of the Odds Ratio, OR.

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 not 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, $OR_{\text{case-control}} \approx RR$

Appendices

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 (some texts do use term 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, \dots, 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 $(E_i \text{ and } E_j) = \text{empty set}$ for all $i \neq j$;
 - (iii) The event $(E_1 \text{ or } E_2 \text{ or } \dots \text{ or } E_n)$ is the entire sample space or population.

In words: E_1, E_2, \dots, 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, \dots, E_n partition the sample space such that $P(E_1) = P(E_2) = \dots = P(E_n)$, then:
- (i) $P(E_i) = 1/n$, for all $i=1, \dots, n$. This means that
 - (ii) the events E_1, E_2, \dots, E_n are equally likely.
- 8) For any event E in the sample space: $0 \leq P(E) \leq 1$.
- 9) $P(\text{empty event}) = 0$. The empty event is also called the **null** set.
- 10) $P(\text{sample space}) = P(\text{population}) = 1$.
- 11) $P(E) + P(E^c) = 1$

B. Addition of Probabilities -

- 1) If events A and B are **mutually exclusive**:
 - (i) $P(A \text{ or } B) = P(A) + P(B)$
 - (ii) $P(A \text{ and } B) = 0$
- 2) More generally:

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$$
- 3) If events E_1, \dots, E_n are all **pairwise mutually exclusive**:

$$P(E_1 \text{ or } \dots \text{ or } E_n) = P(E_1) + \dots + P(E_n)$$

C. Conditional Probabilities -

- 1) $P(B|A) = P(A \text{ and } B) / P(A)$
- 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, \dots, 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 ; $i=1, \dots, k$:

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

E. Bayes Rule -

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

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

Nature

Population/
SampleObservation/
DataRelationships/
ModelingAnalysis/
Synthesis

Appendices

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)

$$\begin{aligned}
 &= [\$1](0.50) + [\$5](0.25) + [\$10](0.15) + [\$25](0.10) \\
 &= \$5.75
 \end{aligned}$$

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:

<u>Value of X, x =</u>	<u>P[X = x] =</u>
\$ 1	0.50
\$ 5	0.25
\$10	0.15
\$25	0.10

The random variable X has *statistical expectation* $E[X]=\mu$

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

Example –

In the “likely winnings” example, $\mu = \$5.75$