Unit 3
Probability Basics

"Chance favours only those who know how to court her"
- Charles Nicolle

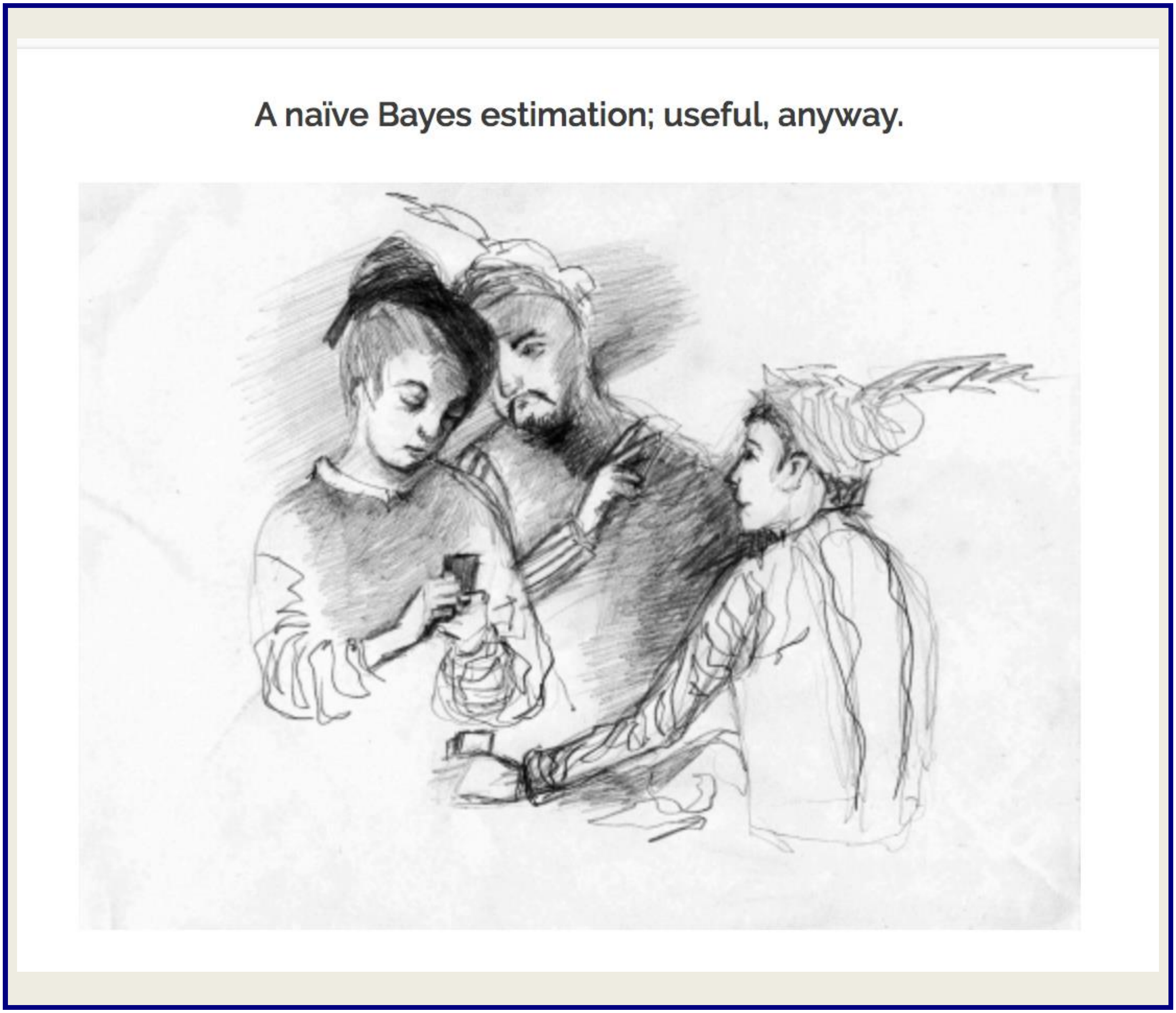
https://lynnandtonicblog.com/tag/cartoons-playing-poker/

A weather report statement such as ***"the chance of rain today is 50%"*** is familiar and we think we have an intuition for what it means. But do we? 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.

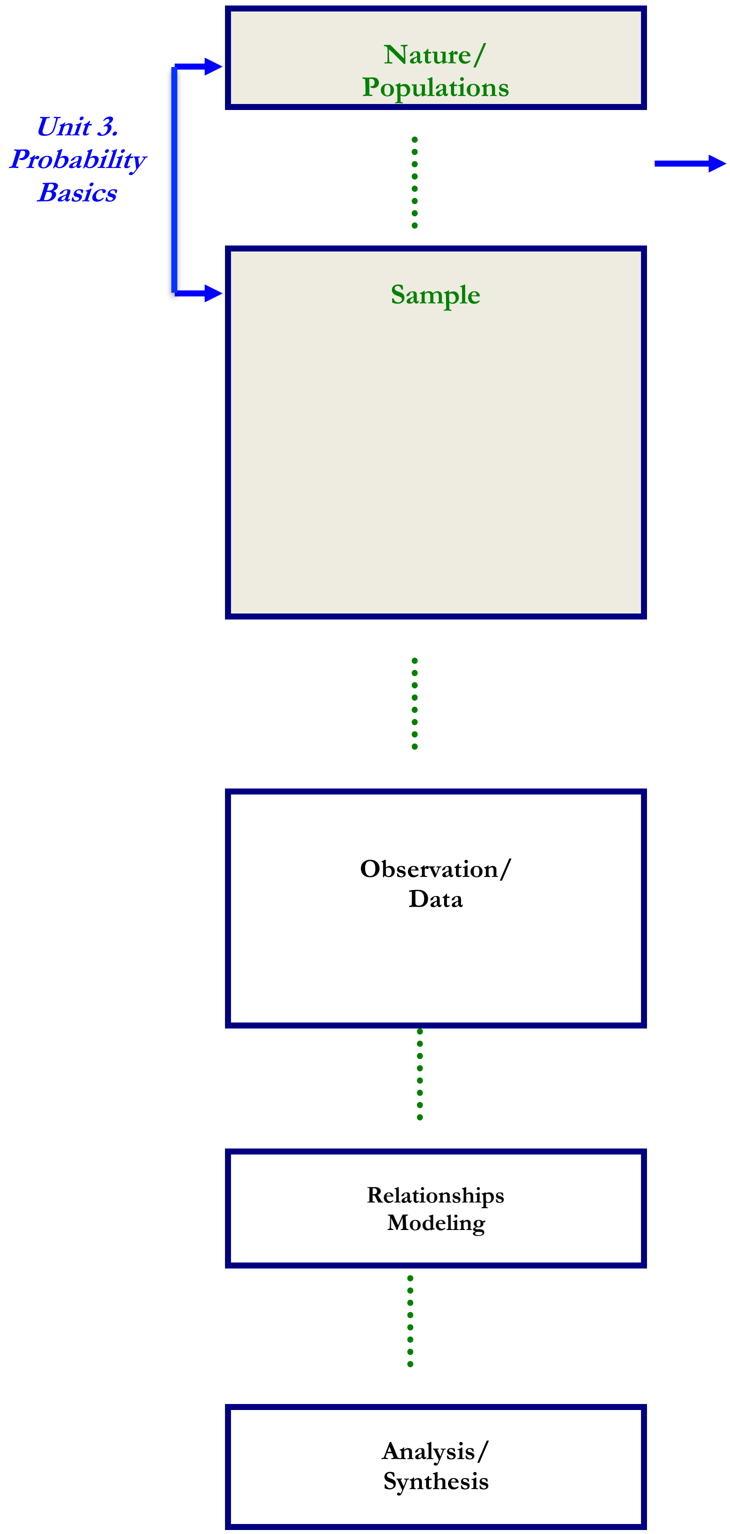
Cheers!



Source: The author is Anastasia Mandel. Downloaded, with permission, from CAUSEweb.org

Table of Contents

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*** of sampling. **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 ("disjoint"), complement, union, intersection.
- Define and explain the distinction between independent and dependent events.
- Explain and use Bayes Rule.

3. Why We Need Probability

In Unit 1 (*Summarizing Data*), our focus 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 recalled that we already 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)*

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

- **Initial lense** - 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 lense** – Suppose now that we are interested in the population that gave rise to the sample. We ask how the sample was obtained. (*Unit 5 – Populations and Samples*).

- **Same enlarged lense** – 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 3 – Probability Basics*).

- Sometimes, the data at hand can be reasonably modeled as random draws from a particular population distribution (this is a model!) – e.g. - Bernoulli, Binomial, Normal. (*Unit 6 – Bernoulli and Binomial Distribution, Unit 7 – 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 8 – Statistical Literacy: Estimation and Hypothesis Testing*).

- **Hypothesis Testing** –We use calculations of probabilities to assess the chances of obtaining statistical summaries such as ours (or more extreme) for a null hypothesis model relative to a competing model called the alternative hypothesis model. (*Units 9-12*)

Going forward:

- We will use the tools of probability in constructing models of what we have observed (*translation: we pretend that our observations are "random draws from some probability distribution model" and run with it.*).

- From there, we will use the tools of probability in confidence interval construction.

- Also from there, we will use the tools of probability in statistical hypothesis testing.

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 outcomes of sampling from some sort of "frame" (more on this later). **Example – A certain population of children at a child care center consists of 53 vaccinated children and 47 unvaccinated children. *Tip* – Populations versus Sample Spaces:** A collection of all possible individuals in nature is termed a "population". The collection of all possible outcomes of sampling the population is termed the "sample space" for that outcome (**E.g – the sample space of all possible samples of size n=10 that could have been sampled from this child care center)**

- Suppose that we can do sampling of the population by the method of **simple random sampling** many times. (more on this in Unit 5, Populations and Samples). **Example continued – If we do ONE random draw from the child care center, what are our chances of selecting a unvaccinated child? Call this our event of interest: "E."**

- As the number of simple random samples draws increases, the proportion of samples in which the event "E" occurs eventually settles down ("in the long run") 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 vaccination status = unvaccinated 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, with simple random sampling, the proportion of samples in which vaccination status=unvaccinated will reach its limiting value of .47**

- The probability of event "E" is the "in the long run" value of the limiting proportion (percent of times).
Example, continued – Probability [E] = .47

Again - This is a "frequentist" approach to probability. There are others:

- **Bayesian** - "Before I do my first toss, I say "This is a fair coin. It lands "heads" with probability ½".

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

- **Subjective** - "This is my lucky coin".

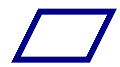
4b. Sample Space, Elementary Outcomes, Event

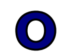
Sample Space:	**"the big pot you are drawing from"**
Elementary Outcome:	**"what you get with ONE draw from your sample space"**
Event:	**"the net result you are interested in; maybe it's a particular collection of elements"**

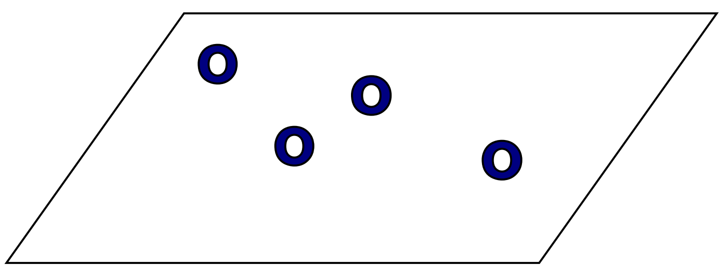
The sample space is comprised of all the possible outcomes of sampling of a population. Each of the possible outcomes in this scenario is also called an "elementary outcome". What we are interested in might be an elementary outcome or it might be some collection of outcomes. Either way, we'll use the term "event" to refer to what is of interest.

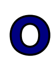
Sample Space -

The **sample space** is defined as the set of all possible (elementary) outcomes of sampling.

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

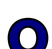
I'm representing an (elementary) outcome with a dark circle that looks like this → 



Elementary Outcome (Sample Point), 

Each possible outcome of the sampling from the sample space is called an **elementary outcome**.

Example –
In this example, the elementary outcome of sampling is a ***pair*** defined as { sex at birth of 1st born twin, sex at birth of 2nd born twin}. There are exactly 4 possibilities: (boy,boy), (boy, girl), (girl, boy), (girl, girl). → Thus, the **sample space** is comprised of these 4 possibilities.
Each of the 4 possible paired outcomes constitutes an **elementary outcome**.

Definition Elementary Outcome:  = {sex at birth of 1st twin, sex at birth of 2nd twin}
Definition Sample Space: All 4 possible elementary outcomes:

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

Event -

Often, our interests are a bit broader than the elementary outcomes. For example, we want a particular "net result"! For example, maybe the "net result" that is desired is that our sampling yields a sex at birth = boy. Snag. **"Boy"** isn't an elementary outcome. But it is a realizable event! In particular, any of 3 elementary outcomes (remember – these are pairs) yields a boy:

(boy, boy), (boy, girl) and (girl, boy).

Thus, the outcome of a "sex at birth = boy" is an example of an **event** that is satisfied by any of 3 elementary outcomes.

An **event "E"** is a collection of one or more elementary outcomes "O".

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

Example, putting it all together –

In the language of these notes, we now say that we are interested in the ***event* of "sex at birth boy"**. The event of a sex at birth boy is satisfied when any of three elementary outcomes has occurred. The three qualifying elementary outcomes are the following:

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

Example – continued

Here are some more examples of "events", together with their associated collections of qualifying elementary outcomes:

Notation	Event	Qualifying elementary outcomes
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}

4c. "Mutually Exclusive ('disjoint')" and "Statistical Independence" Explained

Mutually Exclusive: **"the two cannot occur at the *same* time"**

Statistical Independence: **"thinking over time, the first does not influence the *subsequent* occurrence of the later 2nd"**

The ideas of **mutually exclusive (disjoint)** and **independence** are different and can be confusing. The ways we work with them are also different.

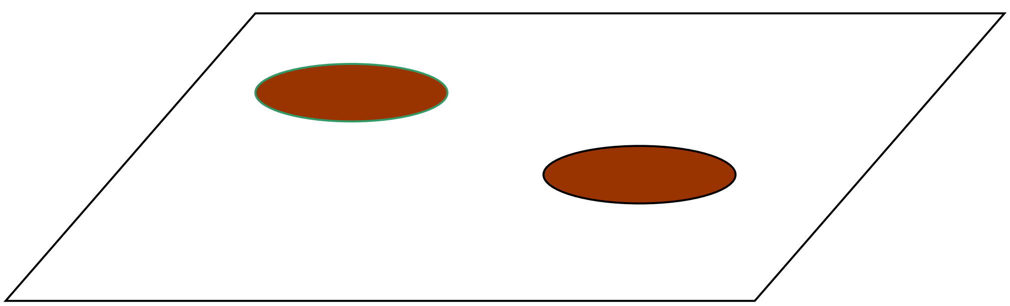
Mutually Exclusive **("disjoint", "cannot occur at the same time")**
Here's a mind numbingly boring example …. Nevertheless, it makes the point.

A **single** coin toss cannot land heads and tails at the same time.

So we say: A coin landed "heads" excludes the possibility that the coin landed "tails".

And then we say: The outcomes "heads" and "tails" in the outcome of a single coin toss are **mutually exclusive**.

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



Some Every day Examples of Mutually Exclusive –

- The weight of a single individual cannot be simultaneously "underweight", "normal", "overweight."
→ These 3 events are mutually exclusive.

- Similarly, the race/Ethnicity of a single individual cannot be simultaneously "White Caucasian", "African American", "Latino", "Native African", "South East Asian", "Other".
→ These 6 events are mutually exclusive.

Statistical Independence ("the occurrence of the first does not change the chances of the second")

It's not a perfect strategy but it's a workable one – **Think about things occurring in two individuals or two things happening over time.** Consider three sets of paired events. Eg.-

- Two individuals. An individual is male and has red hair. The 2 events are "male gender" and "red hair". The forces that determine hair being red might be independent of those that determine gender.

- Over time. An individual is young in age and has high blood pressure. The 2 events are "young age" and "high blood pressure". The biology that produces high blood pressure might be independent of age.

- Over time. First coin toss yields "heads" and second, ***subsequent,*** coin toss yields "tails".
  Paired event = [heads on 1st toss, tails on 2nd toss]
  The influences (wind, trajectory, etc.) that determine the 2nd toss landing heads might be independent of those that determined the 1st toss landing heads.

Two events are statistically independent if the chances, or likelihood/probability, of the second event is in no way related to (dependent on) the chances/likelihood of the first event.

We'll get to calculating chances or likelihoods/probabilities in the next section. Here, consider the game of tossing one coin once and then tossing that same coin a second time. **So what are the chances of two heads?** This is familiar; it is 50% x 50%, representing a 25% chance.

This is a simple example of statistical independence. By the definition of statistical independence (and its applicability to this silly example of a single fair coin):

$$\begin{aligned}\text{"Chances of two heads"} &= \Pr[1^{st}\text{ toss lands heads AND }2^{nd}\text{ toss lands heads}]\\ &= \Pr[1^{st}\text{ toss lands heads}] \times \Pr[2^{nd}\text{ toss lands heads}]\\ &= .50 \times .50\\ &= .25\end{aligned}$$

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 *subsequent* 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) = (1/4), representing a 25% chance.

4d. Complement, Union, Intersection

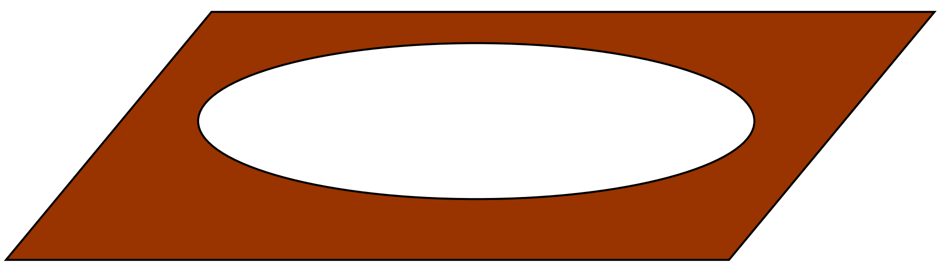
Complement: "The opposite"

Union: "Either A alone occurred, or B alone occurred, or both A and B occurred"

Intersection: "Only what is common to both A and B is what has actually occurred"

Complement ("the opposite occurred") -

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



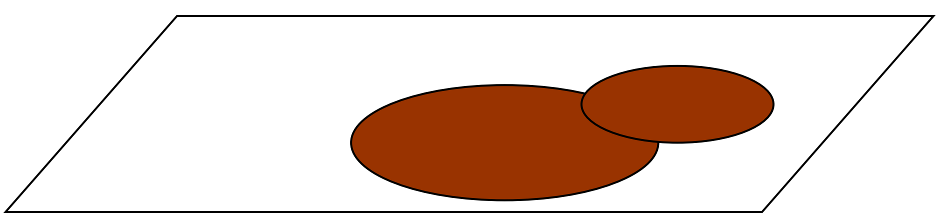
Example –

For the twins example, consider the event E_2 = "two boys" that is introduced on page 9.
The complement of the event E_2 is denoted $E_2{}^c$ and is comprised of 3 elementary outcomes:

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

Union, A or B ("either A occurred OR B occurred OR both A and B occurred") -

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 and sex at birth example, consider the following two events, "A" and "B". In this example, these two events happen to also be elementary outcomes:

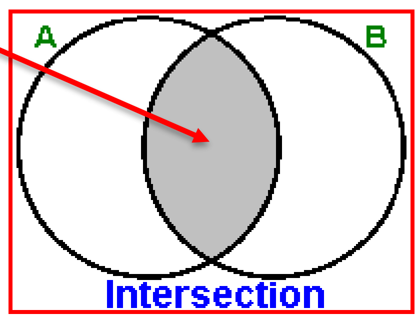
"A" is the event = {boy, girl }.
"B" is the event = {girl, boy }.

The union of events A and B is:
A $\cup$ B = { boy, girl }, { girl, boy }

Intersection, A and B ("only what is common to both A and B occurred") -

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 the joint occurrence of A and B is depicted in the color gray.



Example –

For the twins assigned sex at birth 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. I've highlighted them in bold. The intersection of E_1 and E_2 is " boy and a girl":

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

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_2 is **independent** of the occurrence of event E_1,
>
> THEN - the probability that they BOTH occur is given by multiplying together the two probabilities:
>
> Probability [E_1 occurs ***and*** E_2 occurs]
>
> = Pr[E_1] x Pr[E_2]

Example **(You already know this! The probability of a fair coin landing heads twice is .25) –**

The probability that two tosses of a fair coin both yield heads is a ½ x ½ = ¼ chance.

E_1 = Outcome of heads on 1st toss. Pr[E_1] = .50
E_2 = Outcome of heads on 2nd toss Pr[E_2] = .50

Pr[both land "heads"] = Pr[heads on 1st toss AND heads on 2nd toss]
= Pr[heads on 1st toss] x Pr[heads on 2nd toss *after 1st toss was heads*]
= Pr[heads on 1st toss] x Pr[heads on 2nd toss *anytime because it is independent*]
= Pr[E_1] x Pr[E_2]
= (.50)(.50)
= .25

HOMEWORK DUE Friday October 7, 2022
Question #1 of 3

Let A and B denote two independent genetic traits. Suppose the probability that an individual will exhibit trait A is ½ and the probability that an individual will exhibit trait B is ¾. What is the probability that an individual will exhibit

(a) Both traits?

(b) Neither trait?

(c) Trait A but not trait B?

(d) Trait B but not trait A?

(e) Exactly one trait?

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**.

For now, we consider the scenario in which every elementary outcome in the sample space is equally likely -

- Consider again the **population** of 100 child care children comprised of 53 vaccinated children and 47 unvaccinated children.

- We select one child from this population by the method of simple random sampling **(Imagine. All the children are put into a big pot. The pot is stirred. We reach into the pot and select one).** Each of the 100 children has the same (1 in 100) chance of being selected. **"Simple random sampling"** here means that each of the 100 children has the same (1 in 100) chance of being selected. → **Equally likely** refers to the common 1 in 100 chance of selection that each child has.

- **What if we want to know the chances of selecting a vaccinated child? Or, the chances of selecting an unvaccinated child?**
  This is our **event** of interest. We can define a **random variable** from this event. To do this:

  1) List out all the possible values of the random variable. Answer – vaccinated, unvaccinated.

  2) For each, identify all the "qualifying" equally likely elementary outcomes. Answer - Easy in this example. They are the same.

  3) The probability of each random variable value is obtained by adding together the probabilities of the associated elementary outcomes.

  Example – continued. We'll call this new random variable X.

  X = vaccination status of selected individual child.

And, rather than having to write out "vaccinated" and "unvaccinated", we'll use "0" and "1" as their representations, respectively. Thus, for convenience, we will say

X = 0 when the child selected is "vaccinated"
X = 1 when the child selected is "unvaccinated"

- From an initial collection of equally likely **elementary outcomes** **(each of 100 children has an equal, 1 in 100, chance of being selected)**, together with a defined **event** of interest **(the child selected is vaccinated)** , we can now define what is called a **discrete random variable**. And we can state its **probability distribution**. The probability distribution of a discrete random variable is defined by a listing of all the possible outcomes (this is the **sample space**), together with the associated chances of occurrence (this is the **probability model**).

Discrete Random Variable X = Vaccination Status of Sampled Child

Sample Space **List of all the possible outcomes**	**Probability Model** **List of all the associated chances of occurrence**
0 = vaccinated 1 = unvaccinated ↑ **Tip!** Be sure to check that this listing of all possible outcomes is truly "exhaustive".	0.53 0.47 ↑ Be sure to check that all these probabilities add up to 100% or a total of 1.00.

Some More Formal Language

1. **For a discrete random variable** **(we'll get to continuous random variables later)**, a ***probability model*** is the set of assumptions used to assign probabilities to each possible value (realization) of the random variable. The ***sample space for the discrete random variable*** is the collection, of all the possible values that that random variable can have.

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.

4. When the values of a random variable are all equally likely, the probability model is called a ***uniform probability model.***

Example #1 of a Uniform Probability Distribution – Role ONE die exactly ONE time

WHAT IS A DIE?

In case you are unfamiliar with what is meant by a "die"!

A "die" is a 6-sided cube looking thing. Each side has a number of dots on it. One side has 1 dot, the second side has 2 dots, and so on … the sixth side has 6 dots.

You shake it, toss it in the air, let it land and notice how many dots are facing up.

Source: minneapolisviolinlessons.com

Random Variable X = "face" of die

Sample Space List of all the possible outcomes	**Probability Model** List of all the associated chances of occurrence
1 2 3 4 5 6	1/6 1/6 1/6 1/6 1/6 1/6
Check! This listing of all possible outcomes is "exhaustive".	**Check!** These probabilities total 1 or 100%.

Example #2 of a Uniform Probability Distribution – Toss ONE fair coin exactly ONE time

Random Variable X = "face" of coin

Sample Space List of all the possible outcomes	**Probability Model** List of all the associated chances of occurrence
heads tails **Check!** This listing of all possible outcomes is "exhaustive".	1/2 1/2 **Check!** These probabilities total 1 or 100%.

Example #3 of a Uniform Probability Distribution *(Sampling With Replacement)* – Select 1 digit, by simple random sampling from the list: 1, 2, 3. Return it. Then select a 2nd digit, by equal random sampling from the same list: 1, 2, 3.. N = size of list. n=size of sample.

Random Variable X = {1st selected digit, 2nd selected digit}

Sample Space (Size is 3^2 or N^n) List of all the possible outcomes	**Probability Model** List of all the associated chances of occurrence
1,1 1,2 1,3 2,1 2,2 2,3 3,1 3,2 3,3 **Check!** This listing of all possible outcomes is "exhaustive".	1/9 1/9 1/9 1/9 1/9 1/9 1/9 1/9 1/9 **Check!** These probabilities total 1 or 100%.

Example #4 of a Uniform Probability Distribution *(Sampling Without Replacement)* – Select 1 digit, by simple random sampling from the list: 1, 2, 3. Do *not* return it. Then select a 2nd digit, by equal random sampling from the remaining 2 numbers in the list. N = size of list. n=size of sample.

Random Variable X = {1st selected digit, 2nd selected digit}

Sample Space (Size is 3x2x1 or N x (N-1) x … x (N-n+1) **List of all the possible outcomes**	**Probability Model** **List of all the associated chances of occurrence**
1,2 1,3 2,1 2,3 3,1 3,2 **Check!** This listing of all possible outcomes is "exhaustive".	1/6 1/6 1/6 1/6 1/6 1/6 **Check!** These probabilities total 1 or 100%.

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

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

Now you know how to do this: – The solution is to add up the probabilities of the qualifying elementary outcomes.

Example – Toss one coin two times. Assume that, for each toss, the coin lands heads or tails with the same probability = .50. Assume also that the outcomes of the two tosses are independent (the outcome of the 2nd toss is not influenced or changed in any way depending on the outcome of the 1st toss). Each elementary outcome is a pair = (1st coin face, 2nd coin face)

Event , E	Qualifying Elementary Outcomes, O	Probability (Event, 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

6. The Addition Rule

When to Use the Addition Rule

Tip – Think "one point in time" and think "or"

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

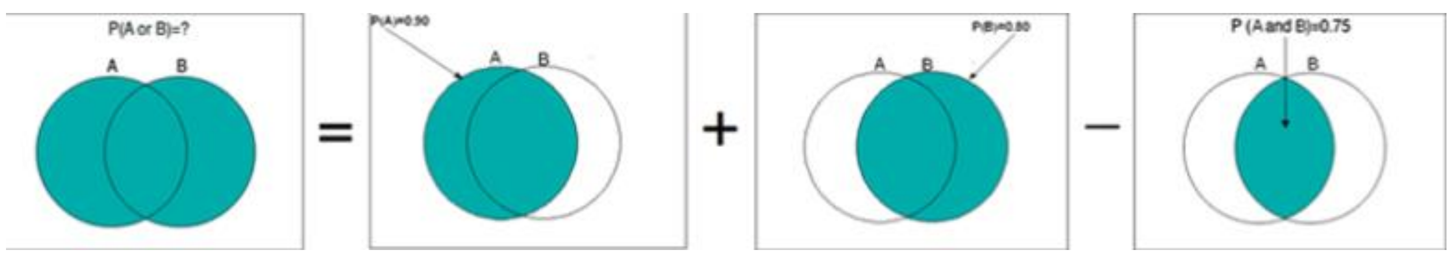
Example 1 – Starbucks will give you a latte with probability = .15. Suppose that, independent of Starbucks, Dunkin Donuts will give you a donut with probability = .35.

Event A: "A" is the event of getting a latte at Starbucks
Thus, Pr [A] = 0.15

Event B: "B " is the event of getting a donut at Dunkin Donuts
Thus, Pr[B] = 0.35

So, what are the chances that you will get a latte ("A") <u>OR</u> a donut ("B")? Reminder - think "either or both" here. Let's call this event "E"

We can depict this using a convenient schematic produced by someone at the University of Florida



Event A *or* B The probability of getting a **latte OR a donut** is represented by the shaded in blue.	**Event A** The probability of getting a **latte** is represented by the shaded in blue. Note we could get either Latte ALONE or Latte + donut	**Event B** The probability of getting a **donut** is represented by the shaded in blue. Note we could get either donut ALONE or donut + Latte	**Event A *and* B** The probability of getting a **latte + donut** is represented by the shaded in blue. **Why the subtraction!!!!** Look left. We've counted (Latte + donut) 2x

Pr [latte OR donut] = Pr [latte] + Pr [donut] - Pr [latte AND donut]

= .15 .35 (.15)(.35)

= **.4475**

What is the probability of the event "latte or donut"?

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

You get the same answer if you add up the probabilities of all the mutually exclusive elementary outcomes that satisfy the occurrence of the event E = "latte OR donut"

Preliminary – List out the entire sample space of this morning's drink and food experience (the elementary outcomes) under the assumption of independence. Note – this means that Pr [NO latte] = .85 and Pr[NO donut] = .65

Sampling Space of Elementary Outcome, O	Probability of O, Pr[O]
{no latte, no donut}	Pr[no latte AND no donut] = (.85)(.65) = .5525
{latte, no donut}	Pr[latte AND no donut] = (.15)(.65) = .0975
{no latte, donut}	Pr[no latte AND donut] = (.85)(.35) = .2975
{latte, donut}	Pr[latte AND donut] = (.15)(.35) = .0525
	Sum = 1 or 100%

Lovely!. So now, what is the probability of the event = "latte or donut"? Easy! The answer is the sum of 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}	.0525
		Brute force sum = **.4475** ***matches***

So why is there a subtraction in the formula

Pr[latte OR donut] = Pr[latte] + Pr[donut] - Pr[BOTH latte and donut]?

Why is this subtraction done?

Answer – The subtraction is done because the elementary outcome {latte, donut}, the portion of the picture that is shaded in gray, appears one time too many. The elementary outcome (latte, donut) appears in the event A and the event (latte,donut) appears in the event B. So, 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}	.0525
		Pr [A] = sum = .15

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

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

Pr[latte OR donut] = Pr[latte] + Pr[donut] - Pr[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 and B] disappears because mutually exclusive events can never both happen at the same time so that probability is zero (Pr [A and B]=0)

$$\Pr[A \cup B] = \Pr[A] + \Pr[B], \textit{ in the special setting where } \Pr[A \text{ and } B] = 0$$

7. The Multiplication Rule – The Basics

When to Use the Multiplication Rule
Tip – Think "two points in time, a first and a second" and 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 13, (4d. The Joint Probability of Independent Events).

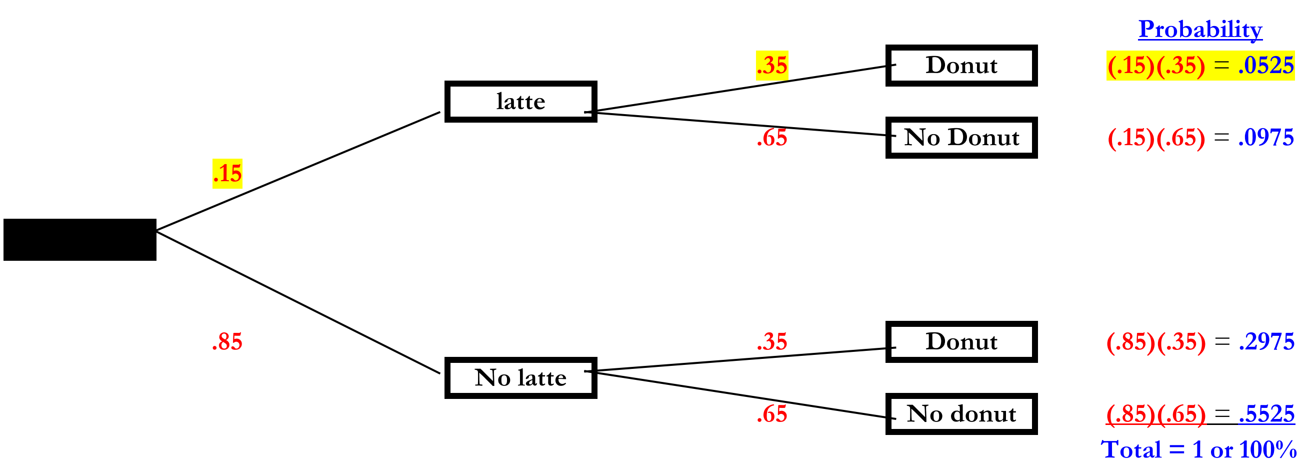
The Joint Probability of Two Independent Events is Obtained by Simple Multiplication
Consider again the example from page 22 – Lattes and Donuts ….– Wouldn't you really rather have both a drink ***and*** something to eat? So now, instead of the event of "latte OR donut", suppose we are interested in the event of **"latte AND donut"**. Yum. We use the multiplication rule to obtain the probability of both (much like we did on page 14 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[**latte and donut**] = Pr[latte] x Pr[donut] = (.15)(.35) = .0525

Introduction to Probability Trees
Lattes and Donuts, continued –
We can construct a little probability tree that shows all the possibilities and their associated probabilities:



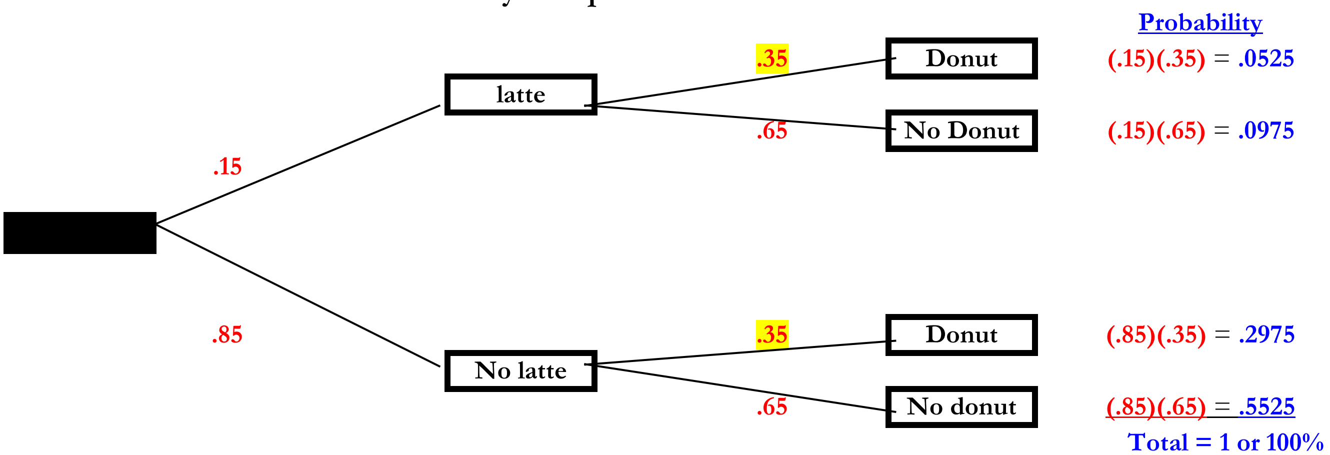
Thus, we see that the Pr[latte and donut] = .15 x .35 = .0525 while
Pr[latte and no donut] = .15 x .65 = .0975, and so on.

8. Conditional Probability

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

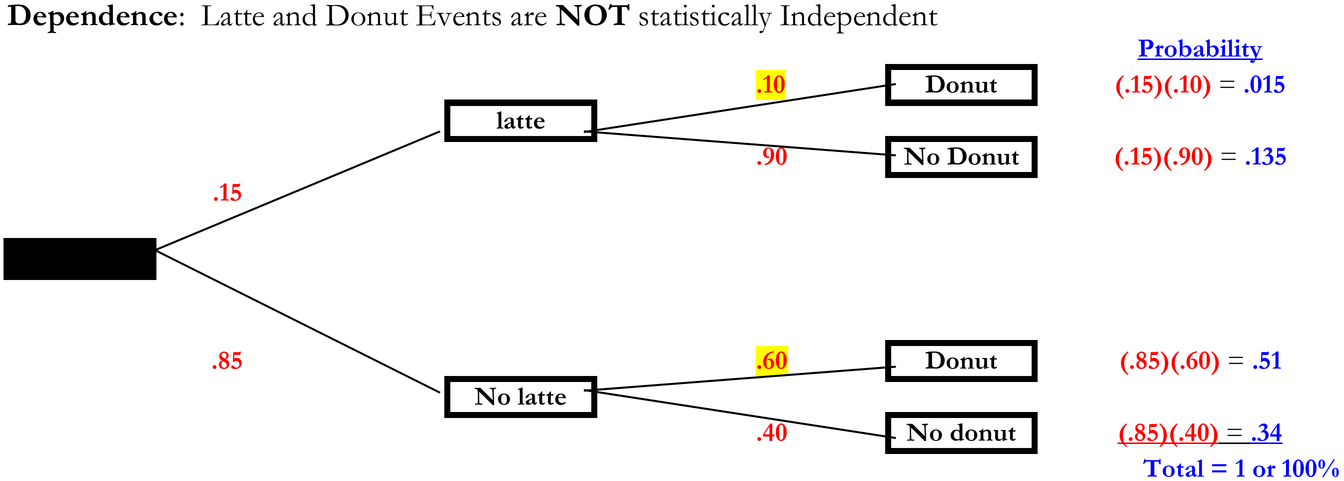
Statistical independence - The occurence of Event A has no influence on the occurrence of Event B
In the "Latte, donut" example, the winning of a donut at Dunkin Donuts was not influenced in any way by the prior winning of a latte at Starbucks. You can see this in the probability tree. Regardless of a winning or not winning of a latte, the chances of a donut from Dunkin Donuts was the same .35 chance

Latte and Donut Events are **Statistically Independent**



Dependence - The occurrence (or non-occurrence) of Event A DOES change the probability of occurrence of Event B
Now suppose that the Dunkin Donut policy on awarding donuts is different for the Starbuck's latte winner than for the Starbucks latte loser. Not very nice, but okay. Suppose that Starbuck latte winners have only a 10% chance of winning a donut while Starbuck's 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



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 event. Not independent! → If two events are dependent, then:

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

It is still possible to calculate the joint probability of both "A" and "B occurring, P(A and B), but this requires a slightly different machinery. In particular, it requires knowing how to relate P(A and B) to what are called **conditional probabilities**. This is explained next.

Conditional Probability

Tip. Imagine you are in a situation where event "A" has occurred. Now you want an answer to the question "Given that A has occurred, now what is the probability that B will occur? *With dependence, the particulars of the event "A" having occurred changes the probability of the event "B"* .

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

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

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

Example - Lattes and Donuts, again. But this time, Dunkin Donuts changes the "rules" depending on the outcome at Starbucks. Using the "tree" on the bottom of page 27 that illustrates "dependence":

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**

- ♣ Lovely. But you want to know your chances of obtaining both something to drink (latte) and something to eat (donut). Pr[latte AND donut]= Pr (A and B). Answer – look back at the formula on the previous page and do a little algebra: multiply both sides by P(A). You get the following.

- ♣ $P(B|A) = \dfrac{P(A \text{ and } B)}{P(A)}$, after multiplying both sides by P(A), gives us what we need:

$$P(A \text{ and } B) = P(A)\, P(B|A)$$ **This is a conditional**

- ♣ Probability of obtaining both latte and donut

= P(A and B) **This is the probability of a latte AND a donut**

= P(A) P(B | A) **This is a pr[latte] x pr[donut given that you have already won latte]**

= 0.15 x 0.10 **0.15 x 0.10**

= 0.015 **Thus, you have a 1.5% chance of both ….**

8b. The Multiplication Rule – A Little More Advanced
Tip – Think "and then"

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

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

provided

(1) Pr[A] is not zero; and

(2) the conditional probabilities are known; and

(3) the conditional probabilities are not zero.

HOMEWORK DUE Friday October 7, 2022
Question #2 of 3

Suppose you are told that pr (right eye is blue) = 1/3 and pr(left eye is blue) = 1/3.

Now suppose I tell you that you may assume that a person's two eyes are ALWAYS the same color

Confirm what you know by intuition, namely that pr(person is blue eyed) = 1/3. To do this, you need to solve pr(blue right eye and blue left eye) = ??

A Useful 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-year 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 need to use to get the answer
A = surviving 1 year	Pr [A] = .75
B = surviving 2 years	Pr [A] = .75 and Pr [B | A] = .85
C = surviving 5 years	Pr [A] = .75 , Pr [B | A] = .85 and Pr [C | B] = .95

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

Pr [survival is 2 years *or more*]
= Pr[surviving 1 year] * Pr[surviving 2 years given survival to 1 year] *** is the same as "times"**
= Pr[A] * Pr[B | A]
= (.75) * (.85)
= **.6375**

Overall 5-year survival = Pr [C] = Pr [B] Pr [C | B]

= Pr [A] Pr [B | A] Pr [C | B]

Pr [survival is 5 years or more]
= Pr[surviving 1 year] * Pr[surviving 2 years given survival to 1 year] * Pr[surviving 5 yrs given 2 yrs]
= Pr[A] * Pr[B | A] * Pr[C | B]
= (.75) * (.85) * (.95)
= **.605625**

<u>So why is this useful?</u> Answer - In BIOSTATS 640, *Intermediate Biostatistics*, we will see that this is **Kaplan-Meier estimation**.

8c. TOOL - The Theorem of Total Probabilities

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

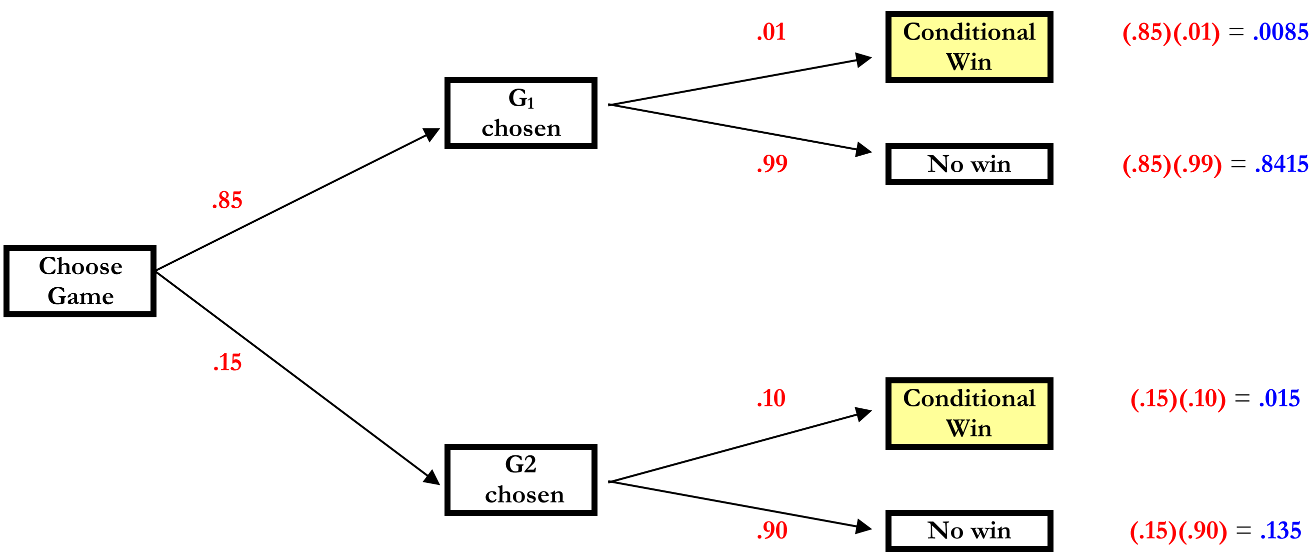
The theorem of total probabilities is used when there is **more than one path to a win** and you want to calculate the **overall probability of a win**.

Chronology over Time -

Step 1: First, 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: Now that you've chosen a game ("given that you choose a game, G_1 or G_2) :
G_1 yields "win" with conditional probability $P(\text{win} \mid G_1) = 0.01$
G_2 yields "win" with conditional probability $P(\text{win} \mid G_2) = 0.10$

Probability Tree Tool -



What is the overall probability of a win, Pr(win)? **Answer – ADD together the qualifying pathways!**

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

What you've just seen is an illustration of the theorem of total probabilities (admittendly, perhaps easier to follow than the statement on the next page).

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

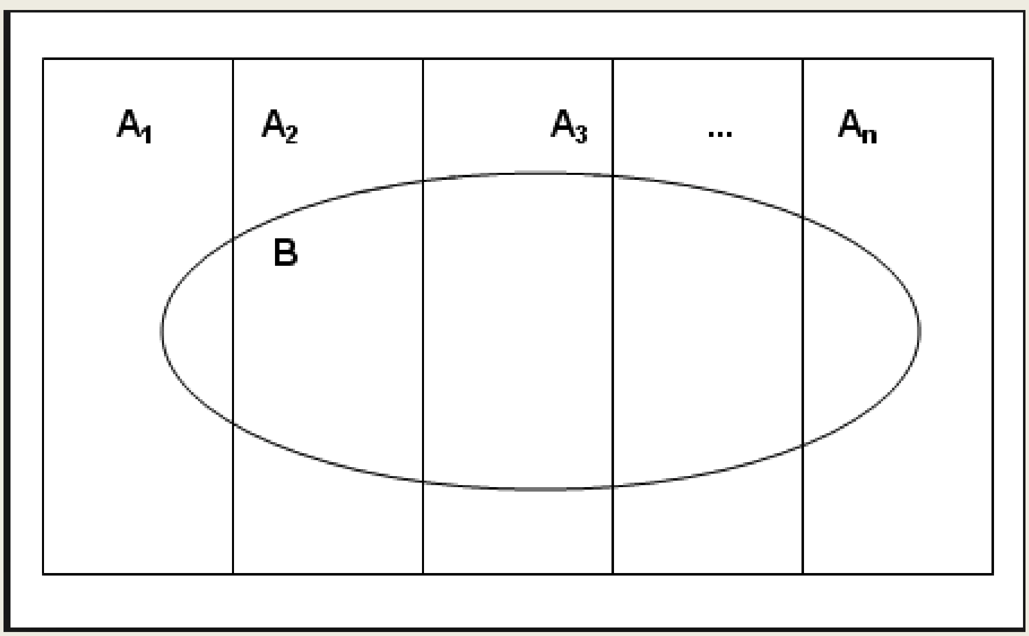
$$S = A_1 \cup A_2 \cup \ldots \cup A_n$$

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

$$P[B] = \underbrace{P[A_1]P[B|A_1]}_{\text{1st way to B}} + \underbrace{P[A_2]P[B|A_2]}_{\text{2nd way to B}} + \ldots + \underbrace{P[A_n]P[B|A_n]}_{\text{last (nth) way to B}}$$

provided the conditional probabilities are known.

Here's a schematic



Source: MathWiki

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 is our tool for computing a joint probability**
   P(A 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 pr[event]**
   $P[E] = P[G_1]P[E|G_1] + P[G_2]P[E|G_2] + \ldots + P[G_K]P[E|G_K]$

Here is the rule. If it looks horribly nasty, hang on. The next page is a "walk-through"

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

$$S = A_1 \cup A_2 \cup \ldots \cup A_n$$

If you are interested in calculating $P(A_i \mid E)$, this is calculated

$$P[A_i|E] = \frac{P(E|A_i)P(A_i)}{P[A_1]P[E|A_1] + P[A_2]P[E|A_2] + \ldots + P[A_n]P[E|A_n]}$$

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

→ **GOAL:** We want to calculate is Pr [Breast Cancer | positive screen] = 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 Bayes Rule solution works …

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

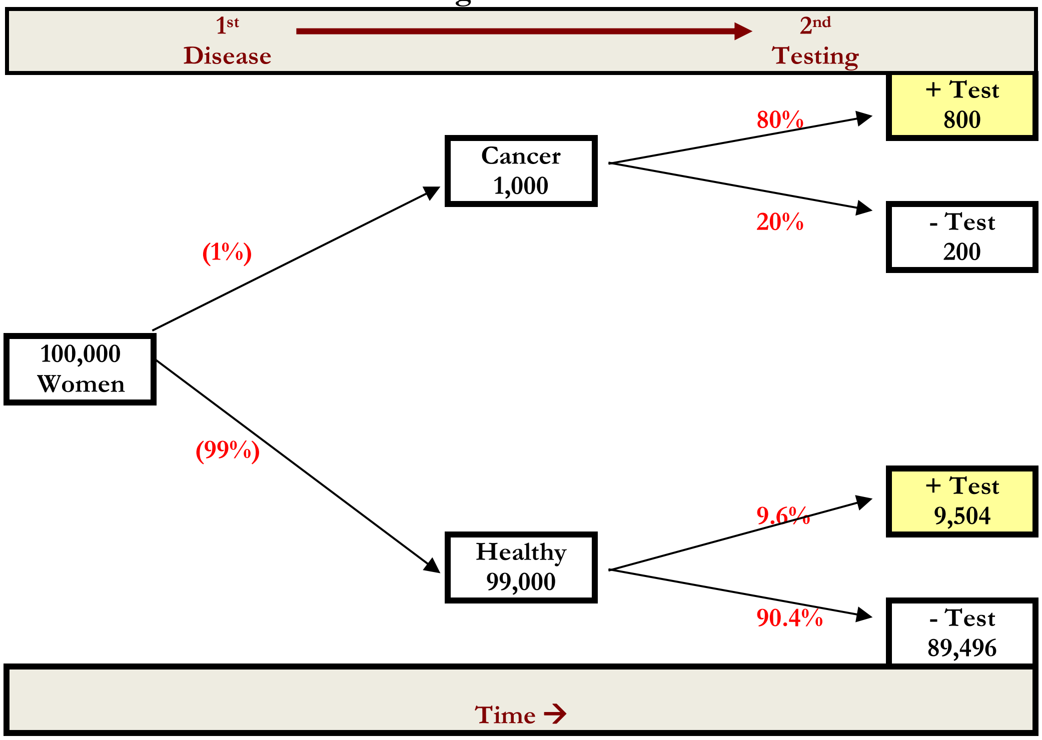
$$= \frac{\Pr(X \mid A)\Pr(A)}{\Pr(X)}$$ because we can re-write the numerator this way

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

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

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

$$\Pr[\text{cancer given positive test}]$$
$$= \Pr[\text{cancer} \mid + \text{test}]$$
$$= \frac{\#\text{ with cancer and +test}}{\#\text{ with +test}}$$
$$= \frac{800}{10{,}304}$$
$$= .078$$

which matches the answer on page 36!

HOMEWORK DUE Friday October 7, 2022
Question #3 of 3

A physician develops a diagnostic test that is positive for 95% of the patients who have disease and is positive for 10% of the patients who do not have disease. Of patients tested, 20% actually have disease.

Suppose you evaluate a patient by administering this diagnostic test and obtain a positive result.

Using the information given, calculate the probability that this patient has disease.

Appendix A.
Some Elementary Laws of Probability

A. Definitions

1) One sample point corresponds to each possible outcome of sampling.

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 outcomes "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 individual 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 A_1, A_2, ..., A_n are said to partition the sample space or population if:

   (i) A_i is contained in the sample space, and
   (ii) The event (A_i and A_j) = empty set for all $i \neq j$;
   (iii) The event (A_1 or A_2 or ... or A_n) is the entire sample space or population."

   In words: A_1, A_2, ..., A_n are said to partition the sample space if they are pairwise mutually exclusive and collectively exhaustive.

Not sure of the meaning of "partitioning the sample space"? Imagine a whole pie that you cut into pieces. You've just partitioned the pie.

A. Definitions - continued

7) If the events $A_1, A_2, ..., A_n$ partition the sample space such that $P(A_1) = P(A_2) = ... P(A_n)$ **(recognize this? "equally likely")** , then:

   (i) $P(A_i) = 1/n$, for all i=1,...,n. This means that

   (ii) the events $A_1, A_2, ..., A_n$ 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 $E_1, \ldots, E_n$ are all pairwise mutually exclusive:
   $P(E_1 \text{ or } \ldots \text{ or } E_n) = P(E_1) + \ldots + P(E_n)$

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 $E_1, \ldots, E_n$ be mutually exclusive events that partition the sample space. The unconditional probability of the event E can then be written as a weighted average of the conditional probabilities of the event E given the A_i; i=1, ..., n:

$$P(E) = P(E|A_1)P(A_1) + P(E|A_2)P(A_2) + \ldots + P(E|A_n)P(A_n)$$

E. Bayes Rule

If the sample space is partitioned into n disjoint events $A_1, \ldots, A_n$, then for any event E:

$$P(A_j|E) = \frac{P(E|A_j)P(A_j)}{P(E|A_1)P(A_1) + P(E|A_2)P(A_2) + \ldots + P(E|A_n)P(A_n)}$$

Appendix B.
Introduction to the Concept of Expected Value

We'll talk about the concept of "expected value" again later. This is just a "warm up".

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.

Statistical Expectation

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* $E[X]=\mu$

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

Example –

In the "likely winnings" example, the statistically expected winning is μ = \$5.75 and is obtained as:

\$5.75 = [\$1 x *50% chance*] + [\$5 x *25% chance*] + [\$10 x *15% chance*] + [\$25 x *10% chance*]

= [\$1 x .50] + [\$5 x .25] + [\$10 x .15] + [\$25 x .10]