TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI

Probability: Multiple Events
MATHematics MAGnification™

Dr. Ralph deLaubenfels

TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI TSI

Teacher-Scholar Institute

Columbus, Ohio

2018

PROBABILITY: MULTIPLE EVENTS MAGNIFICATION

This is one of a series of very short books on math, statistics, and physics called "Math Magnifications." The "magnification" refers to focusing on a particular topic that is pivotal in or emblematic of mathematics.

OUTLINE

This magnification is one of four intuitive introductory magnifications about probability. See [2] for more topics and details, especially extensive examples and homework, or see [1, Chapter I] for an informal introduction to probability.

Given two events A and B, this magnification will discuss (A and B), (A or B), and (A given B), denoted $(A|B)$. Examples include cat and dog ownership, being clean versus being generous, marbles, false accusations, and the Three Stooges.

The only prerequisites for this magnification are arithmetic, including fractions and percents, and a very small amount of algebra, mainly solving linear equations.

INTRODUCTION.

It is often the case that individual events are understood, but relations between those events are not. For example, a cat shelter might keep track of how many people get cats and a dog shelter of how many get dogs, without there being any sharing of information between the shelters. Specifically, we might want to answer the following questions.

1. What percentage of people have both a cat and a dog?
2. What percentage of people have either a cat or a dog?
3. What percentage of people have neither a cat nor a dog?
4. What percentage of people have a cat but no dog?
5. What percentage of people have a dog but no cat?
6. What percentage of people have a cat or a dog but not both?
7. What percentage of cat owners have a dog?
8. What percentage of dog owners have a cat?

Definitions 0.1. Recall, from [1] or [2], that **events** are the sets of outcomes of an experiment that we take probabilities of; denote by $P(A)$ the **probability** of the event A. A **sure** event has probability one, an **impossible** event has probability zero, and, for any event A, $0 \leq P(A) \leq 1$.

The **sample space** or **universe** is the set of all possible outcomes of an experiment; it is surely a sure event.

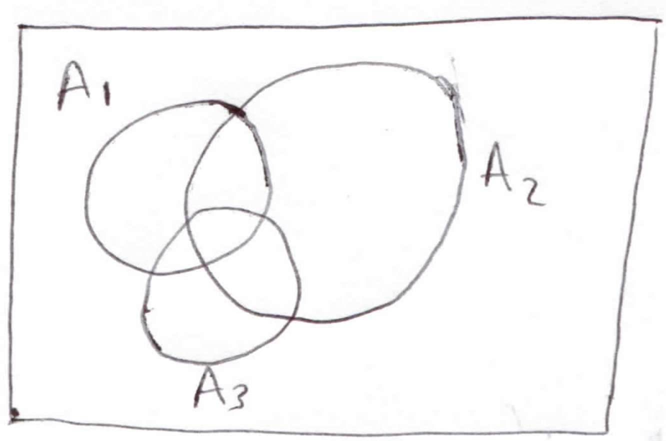
The questions above involve the following two events:

$C \equiv$ a randomly chosen person has a cat; $D \equiv$ a randomly chosen person has a dog.

The experiment is "choose a person at random."

Recall, from [1] or [2], a **Venn diagram**: The universe is a rectangle, events are discs, with probability of an event equal to the area of the disc representing the event.

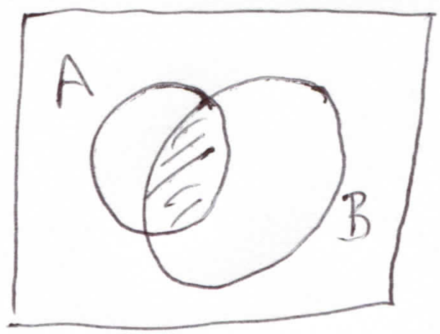
Here is a Venn diagram for three events A_1, A_2, and A_3.



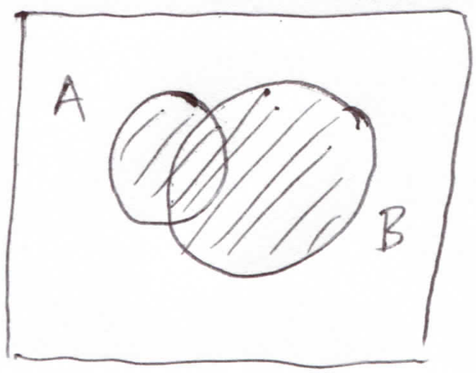
CHAPTER I: Intersections, Unions, and Complements.

Definitions and Terminology 1.1. We need the following vocabulary. Throughout, A and B are events. Each definition is accompanied by its Venn diagram.

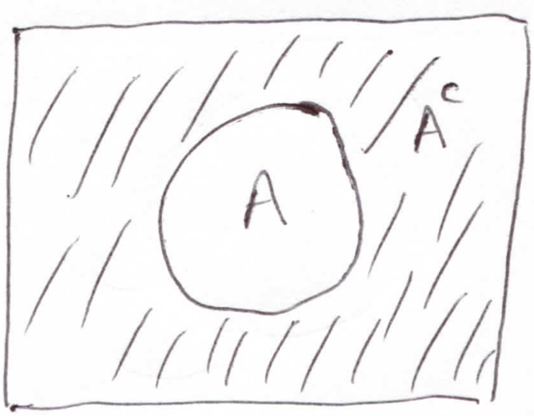
$(A \cap B)$, reading "A intersect B" or "A and B" is the set of all outcomes that are in both A and B.



$(A \cup B)$, reading "A union B" or "A or B" is the set of all outcomes that are either in A or B or both; this is sometimes called *inclusive or*, since it contains (A and B) $= (A \cap B)$.



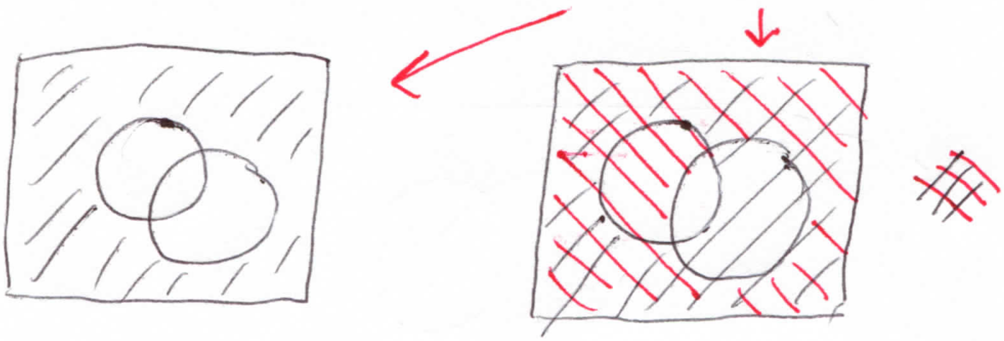
A^c, reading "A complement" or "complement of A" or "not A" is the set of all outcomes that are not in A.



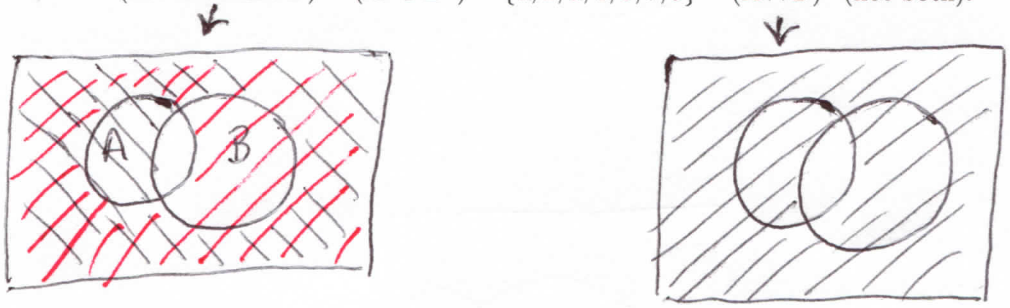
Examples 1.2. (a) Let our sample space be $\{0, 1, 2, \dots, 9\}$, A the even numbers $\{0, 2, 4, 6, 8\}$, and B the numbers greater than 5 $\{6, 7, 8, 9\}$.

Then

$$(A \cup B) = \{0,2,4,6,7,8,9\},\ (A \cap B) = \{6,8\},\ A^c = \{1,3,5,7,9\},\ B^c = \{0,1,2,3,4,5\},$$
$$(\text{neither } A \text{ nor } B) = \{1,3,5\} = (A \cup B)^c = (A^c \cap B^c), \text{ and}$$



$$(\text{not } A \text{ or not } B) = (A^c \cup B^c) = \{1,2,3,4,5,7,9\} = (A \cap B)^c \text{ (not both).}$$



Those equalities $(A \cup B)^c = (A^c \cap B^c)$ and $(A \cap B)^c = (A^c \cup B^c)$ are called **De Morgan's Laws**.

(b) Let C and D be as in the Introduction. Let's translate the events whose probabilities are requested in questions 1–6 of the Introduction.

1. A randomly chosen person has both a cat and a dog: $C \cap D$.

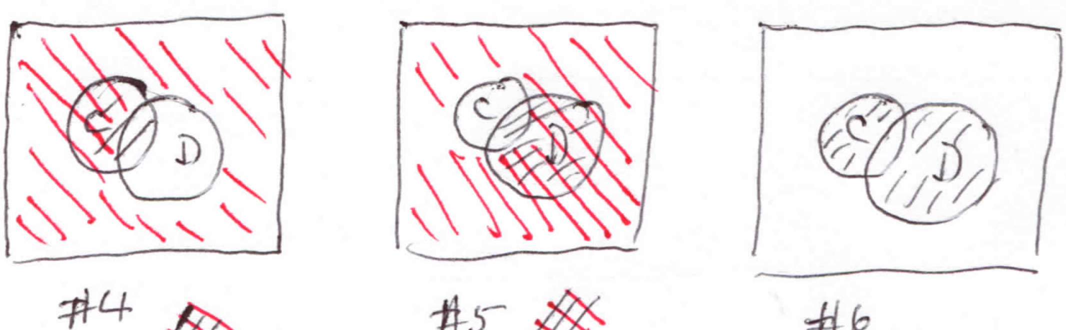
2. A randomly chosen person has either a cat or a dog: $C \cup D$.

3. A randomly ... has neither a cat nor a dog: $(C \cup D)^c$.

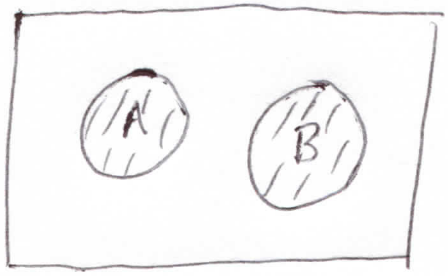
4. ... a cat but no dog: $C \cap D^c$.

5. ... a dog but no cat: $D \cap C^c$.

6. ... a cat or a dog but not both: $[(C \cap D^c) \cup (D \cap C^c)]$ OR $[(C \cup D) \cap (C \cap D)^c]$.



Definition 1.3. A and B are **mutually exclusive** or **disjoint** if $P(A \cap B) = 0$.



Property of Probability 1.4. If $A_1, A_2, A_3, \ldots$ is a sequence of pairwise disjoint events, then

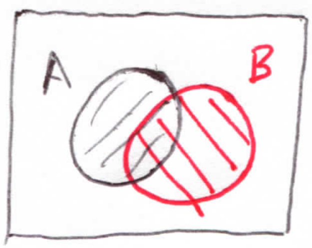
$$P(A_1 \cup A_2 \cup A_3 \ldots) = P(A_1) + P(A_2) + P(A_3) + \ldots$$

For a pair of events that might not be disjoint, we have the following generalization of 1.4.

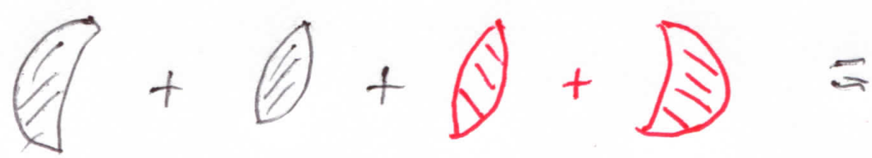
Addition Law 1.5. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

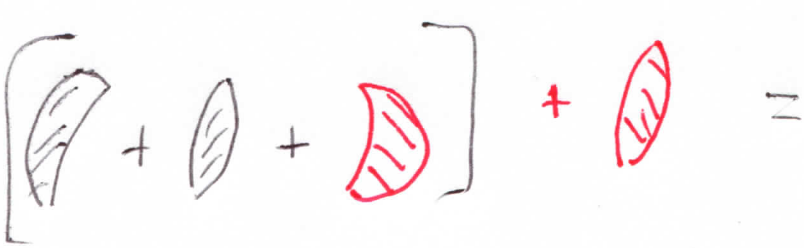
Proof: In words, $P(A \cap B)$ is counted twice when we write down $P(A) + P(B)$, so we correct by subtracting one of those $P(A \cap B)$s.

In pictures:



$P(A) + P(B) =$





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

Law of the Complement 1.6. For any event A, $P(A^c) = 1 - P(A)$.

Proof: Let S be the sample space. Then

$$1 = P(S) = P(A \cup A^c) = P(A) + P(A^c),$$

by 1.4 or 1.5. □

Examples 1.7. Suppose 70% of the population is clean, 80% is generous, and 95% is clean or generous. Find what percent of the population is

(a) clean and generous;

(b) dirty and stingy;

(c) dirty or stingy;

(d) clean and stingy;

(e) dirty and generous;

(f) clean or generous, but not both.

Solutions. These percentages are disguised probabilities. Take the experiment to be "choose a person at random," then define events

$$C \equiv \text{person chosen is clean} \quad \text{and} \quad G \equiv \text{person chosen is generous}.$$

We have

$$P(C) = 0.7, \quad P(G) = 0.8, \quad P(C \cup G) = 0.95.$$

(a) The Addition Law implies that

$$0.95 = 0.7 + 0.8 - P(C \cap G),$$

so that $P(C \cap G) = 0.7 + 0.8 - 0.95 = 0.55$, or 55 percent.

(b) In words, we are avoiding both C and G, thus we want

$$P(C^c \cap G^c) = P\left((C \cup G)^c\right) = 1 - P(C \cup G) = 1 - 0.95 = 0.05,$$

or 5 percent, by the Law of the Complement 1.6.

See the first drawing in Examples 1.2.

(c)

$$P(C^c \cup G^c) = P\left((C \cap G)^c\right) = 1 - P(C \cap G) = 1 - 0.55 = 0.45,$$

or 45 percent, by the Law of the Complement.

See the second drawing in Examples 1.2.

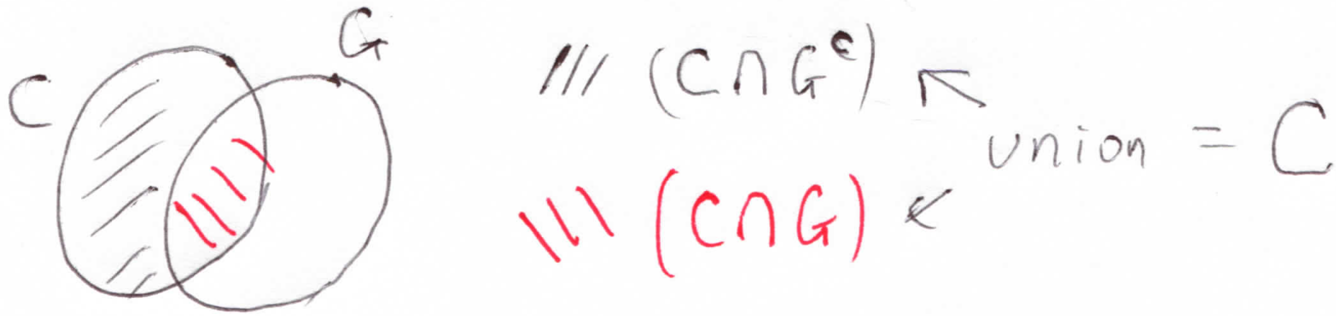
(d) We want $P(C \cap G^c)$. By 1.4 or 1.5,

$$P(C) = P(C \cap G) + P(C \cap G^c),$$

thus

$$P(C \cap G^c) = P(C) - P(C \cap G) = 0.7 - 0.55$$

equals 15 percent.

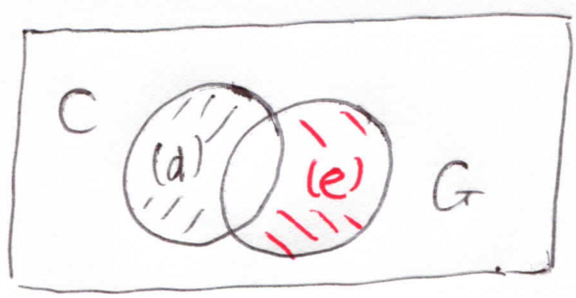


(e) Now we want $P(C^c \cap G)$; as with (d), this equals

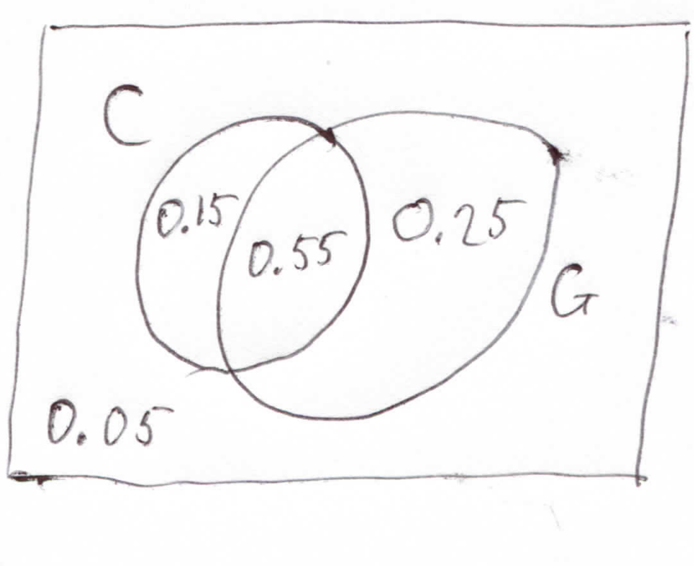
$$P(G) - P(C \cap G) = 0.8 - 0.55$$

equals 25 percent.

(f) This is the disjoint union of (d) and (e): $15 + 25 = 40$ percent.



A more systematic attack would be to solve (a), (b), (d), and (e) first, to give us the following partitioned Venn diagram.



A neater picture of the same data would be the following matrix.

	C	C^c
G	0.55	0.25
G^c	0.15	0.05

In a matrix of this form, each number is the probability of an intersection; the matrix is a compact picture of the following data:

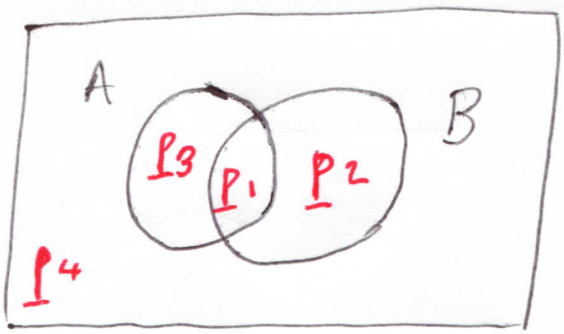
$$P(C \cap G) = 0.55, P(C^c \cap G) = 0.25, P(C \cap G^c) = 0.15, P(C^c \cap G^c) = 0.05.$$

Since these intersections are disjoint events, *any* probability involving C and G is obtained by adding up the needed entries in the matrix; e.g., as in (f),

$$P((C \cap G^c) \cup (C^c \cap G)) = 0.15 + 0.25 = 0.40.$$

Note that the entries in the matrix sum to one.

Remark 1.8. More generally, the Venn diagram



is equivalent to the matrix of probabilities

	A	A^c
B	p_1	p_2
B^c	p_3	p_4

See [2, Chapter III] for more examples of Addition Law and Law of the Complement.

CHAPTER II: Conditional Probability

Conditional probability addresses the fact that probabilities might change when we have more information.

Definition 2.1. If A and B are events, then the (conditional) **probability of A, given B**, denoted $P(A|B)$, is the probability of A under the assumption of B.

Example 2.2. Let A be the event "it will rain in the next hour," and B the event "it is cloudy now."

Even without meteorology degrees, we might believe that

$$P(A|B) > P(A).$$

$P(A)$ is a blind probability; we are sitting in our study with the blinds closed, perhaps armed with data about rain on this day many years in the past. The conditional probability $P(A|B)$ means we open the blinds and look outside; if we see clouds we consider rain more likely.

If we *don't* see clouds when we open the blinds, we would also modify our probability; rain is unlikely when it's not cloudy. This example's choice of A and B implies that

$$P(A|B^c) < P(A) < P(A|B);$$

$P(A)$ can be shown to be a weighted average of $P(A|B)$ and $P(A|B^c)$.

Discussion 2.3. Conditional probability is probably the most important idea in probability. It is implicit in many of our most consequential, commonly accepted statements. Here are some examples.

In court, we look at

$$P(\text{evidence} \mid \text{accused is innocent}) \quad (*),$$

the probability of the evidence under the assumption of the accused's innocence. The smaller this probability is, the less inclined we are to believe the accused is innocent. For example,

$$P(\text{accused was seen dropping something near the littering incident} \mid \text{accused is innocent of littering}),$$

is less than

$$P(\text{accused was seen near the littering incident} \mid \text{accused is innocent of littering}),$$

thus the former evidence "accused was seen dropping something near the littering incident" is more likely than the latter evidence "accused was seen near the littering incident" to make us suspect the accused is guilty of littering.

"Presumption of innocence" means insisting that the probability (*) be *really* small before we convict.

A quick foreshadowing of the intellectual future: in statistical inference, the conditional probability of (*) is the *p-value* for the *null hypothesis* of the accused being innocent. We will talk about this in a future magnification.

The scientific method may be summarized as follows: if

$$P(\text{data} \mid \text{Model One}) > P(\text{data} \mid \text{Model Two}),$$

then we prefer Model One; in words, Model One "fits the data" better than Model Two.

Then there is the expanding world of health advice, both physical and cultural. Consider the statement "smoking causes cancer." Does this mean everyone who smokes gets cancer? It's not true, so I hope that's not what is meant. Does the statement mean everyone who doesn't smoke avoids cancer? This is also not true. What's probably meant is that the probability of cancer *increases* under the influence of smoking: if C is cancer and S is smoking, then our statement probably means

$$P(C|S) > P(C) > P(C|S^c).$$

Example 2.4. Let C and D be the events in the Introduction. We claim that Question 7 in the Introduction is asking for

$$P(D|C),$$

while Question 8 is

$$P(C|D).$$

Here is how you could think of Question 7. Imagine an auditorium containing all cat owners and no one else. It is among this group that you are calculating the relative frequency of dogs; e.g., if the auditorium contained 200 cat owners, among which there were 50 dog owners, the desired probability would be

$$\frac{50}{200} = \frac{1}{4},$$

or 25 percent.

This probability fraction is only among cat owners; it begins with the assumption that all persons considered are cat owners, hence we are "given C."

Question 8 similarly begins by making all people who are not dog owners leave the room; we are talking about cat owners among dog owners only.

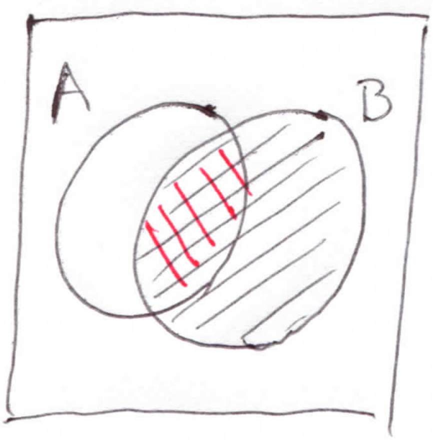
Discussion and Formula 2.5. In our hypothetical numbers in Example 2.4 to describe Question 7 in the Introduction, hypothesize also that there are 1,000 people in our universe. Then we have

$$P(D|C) = \frac{50}{200} = \frac{\frac{50}{1,000}}{\frac{200}{1,000}} = \frac{P(D \cap C)}{P(C)}.$$

In general, for any events A and B,

$$P(A|B) = \frac{P(A \cap B)}{P(B)}.$$

As with any Venn diagram, we are talking about relative area, but we are now in the shrunken universe of B.



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

Examples 2.6. (a) Here is the distribution of 80 marbles, where B stands for blue, R for red, Y for yellow, C for cracked, C^c for not cracked; each entry is the intersection of its row and column; that is, there are 5 blue marbles that are cracked, 6 red marbles that are not cracked, etc.

	B	R	Y
C	5	4	5
C^c	40	6	20

(i) What percent of marbles is cracked?

(ii) What is the probability that a randomly chosen marble is yellow?

(iii) Given that a marble is blue, what is the probability that it is cracked?

(iv) Which color of marble is most likely to be cracked?

(v) If a marble is red, what is the probability it is not cracked?

(vi) If a marble is not cracked, what is the probability it is red?

(vii) What percent of cracked marbles is yellow?

(b) For the probabilities in Examples 1.7, find the following.

(i) the percent of the dirty population that is generous;

(ii) the percent of the stingy population that is dirty;

(iii) if a randomly chosen person is generous, what is the probability that said person is clean;

(iv) given that a randomly chosen person is clean, what is the probability that said person is generous.

Solutions. (a) First, let's change the entries to probabilities by dividing by 80. Each letter is shorthand for "probability that randomly chosen marble is [letter]."

	B	R	Y
C	$5/80 = 0.0625$	$4/80 = 0.05$	$5/80 = 0.0625$
C^c	$40/80 = 0.5$	$6/80 = 0.075$	$20/80 = 0.25$

(i) $0.0625 + 0.05 + 0.0625 = 0.175$

(ii) $0.0625 + 0.25 = 0.3125$

(iii) $P(C|B) = \frac{P(C \cap B)}{P(B)} = \frac{0.0625}{0.0625+0.5} = \frac{1}{9} = 0.111\ldots$

(iv) $P(C|R) = \frac{0.05}{0.05+0.075} = 0.4, P(C|Y) = \frac{0.0625}{0.0625+0.25} = 0.2$; **red** gives the largest conditional probability.

(v) $P(C^c|R) = \frac{0.075}{0.05+0.075} = 0.6$; NOTE that this is $1 - P(C|R)$.

(vi) $P(R|C^c) = \frac{0.075}{0.5+0.075+0.25} = 0.090909\ldots$

(vii) $P(Y|C) = \frac{0.0625}{0.0625+0.05+0.0625} \sim 0.3571$

(b) Use the matrix at the end of the solutions to Examples 1.7.

(i) $P(G|C^c) = \frac{P(G \cap C^c)}{P(C^c)} = \frac{0.25}{0.25+0.05} = \frac{5}{6} = 0.8333\ldots$

(ii) $P(C^c|G^c) = \frac{P(C^c \cap G^c)}{P(G^c)} = \frac{0.05}{0.15+0.05} = \frac{1}{4} = 0.25$

(iii) $P(C|G) = \frac{0.55}{0.55+0.25} = 0.6875$

(iv) $P(G|C) = \frac{0.55}{0.55+0.15} = \frac{55}{70} \sim 0.7857$

Formula 2.5 immediately gives us the following formula for intersections of events.

Multiplication Law 2.7. For any events A, B,

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

Examples 2.8. (a) Suppose two cards are dealt at random from a 52-card pack containing four aces. What is the probability that both cards are aces?

(b) Suppose, very hypothetically, the following, where "drug" means "illegal drug."

Eighty percent of drug users sweat profusely, ten percent of people who don't use drugs sweat profusely, and two percent of the population uses drugs.

The following might seem intuitively reasonable from the data just given:

DRUG TEST. If someone sweats profusely, said person will be accused of using drugs.

What percent of accusations of drug use are false?

(c) Moe, Larry, and Curly are camping.

Forty percent of tents are pitched by Moe, fifty-nine percent by Larry, and one percent by Curly. Fifteen percent of tents pitched by Moe fall, ten percent of tents pitched by Larry fall, and ninety percent of tents pitched by Curly fall.

If a tent falls, what it is the probability it was pitched by Curly? (The previous paragraph might incline us to blaming Curly whenever a tent falls.)

Solutions. (a) Let $A \equiv$ "first card is ace," $B \equiv$ "second card is ace."

We want $P(A \cap B) = P(A|B)P(B)$. Neither of those probabilities in that product look obvious. But notice now that

$$P(A \cap B) = P(B \cap A) = P(B|A)P(A) = \left(\frac{3}{51}\right)\left(\frac{4}{52}\right) = \left(\frac{3 \times 4}{51 \times 52}\right) = \frac{1}{(17 \times 13)} = \frac{1}{221}.$$

(b) Let's translate probabilities. Denote S for (randomly chosen person) sweats, D for (randomly chosen person) takes drugs.

We have

$$0.02 = P(D),\ 0.8 = P(S|D),\ 0.1 = P(S|D^c),$$

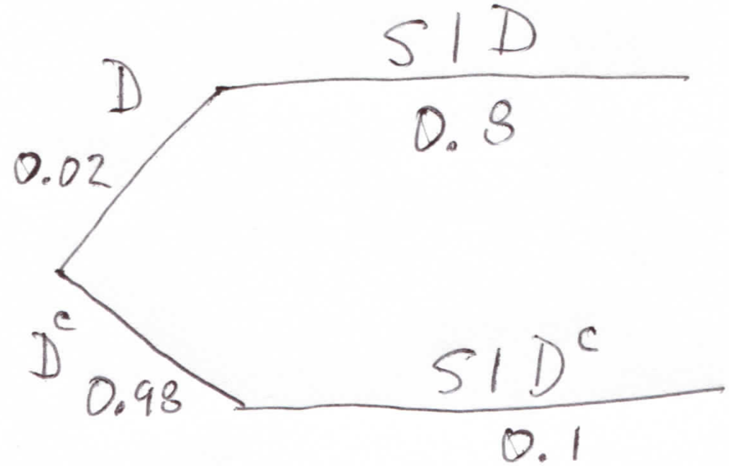
and we want

$$P(D^c|S).$$

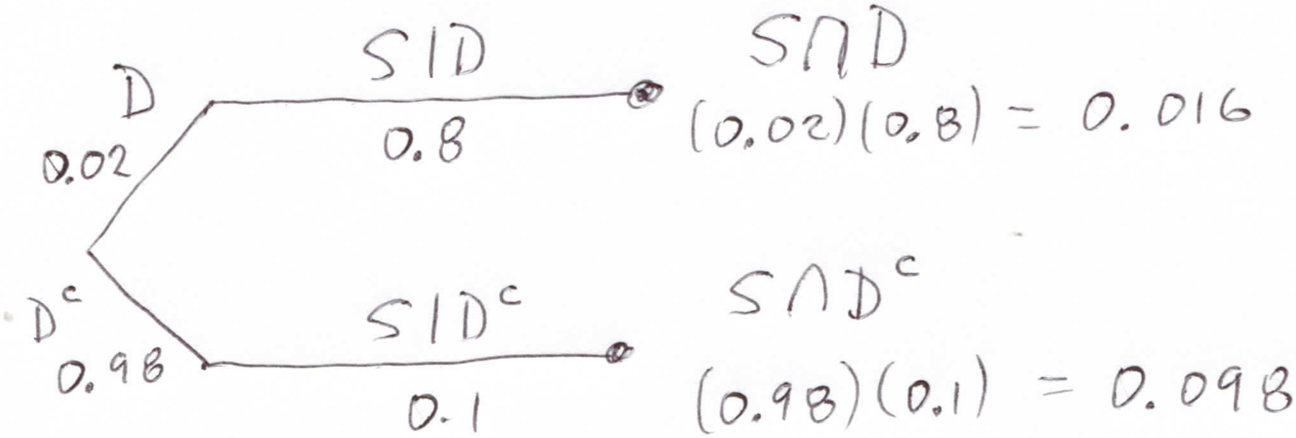
For lack of anything better to do, we might as well translate what we want with Formula 2.5:

$$\text{GET } \frac{P(D^c \cap S)}{P(S)}.$$

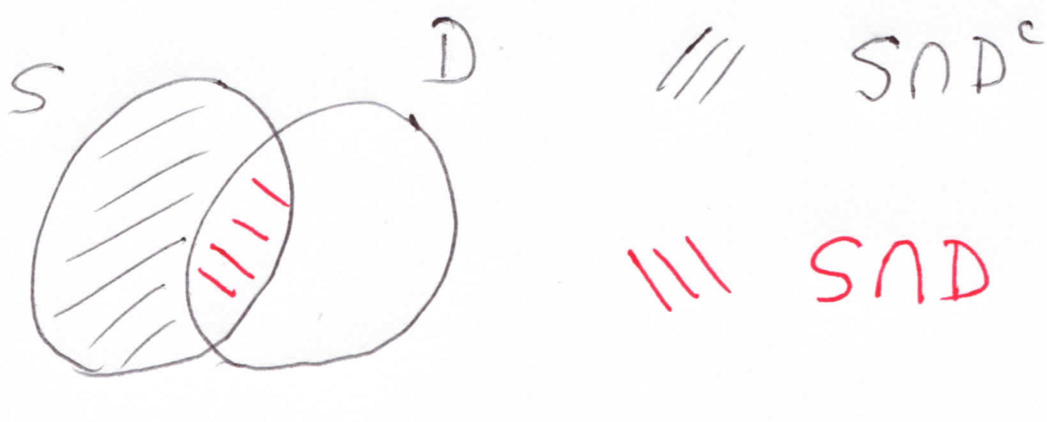
The following "tree diagram" is useful for confusing mixtures of conditional probabilities; we used the Law of the Complement for $P(D^c)$.



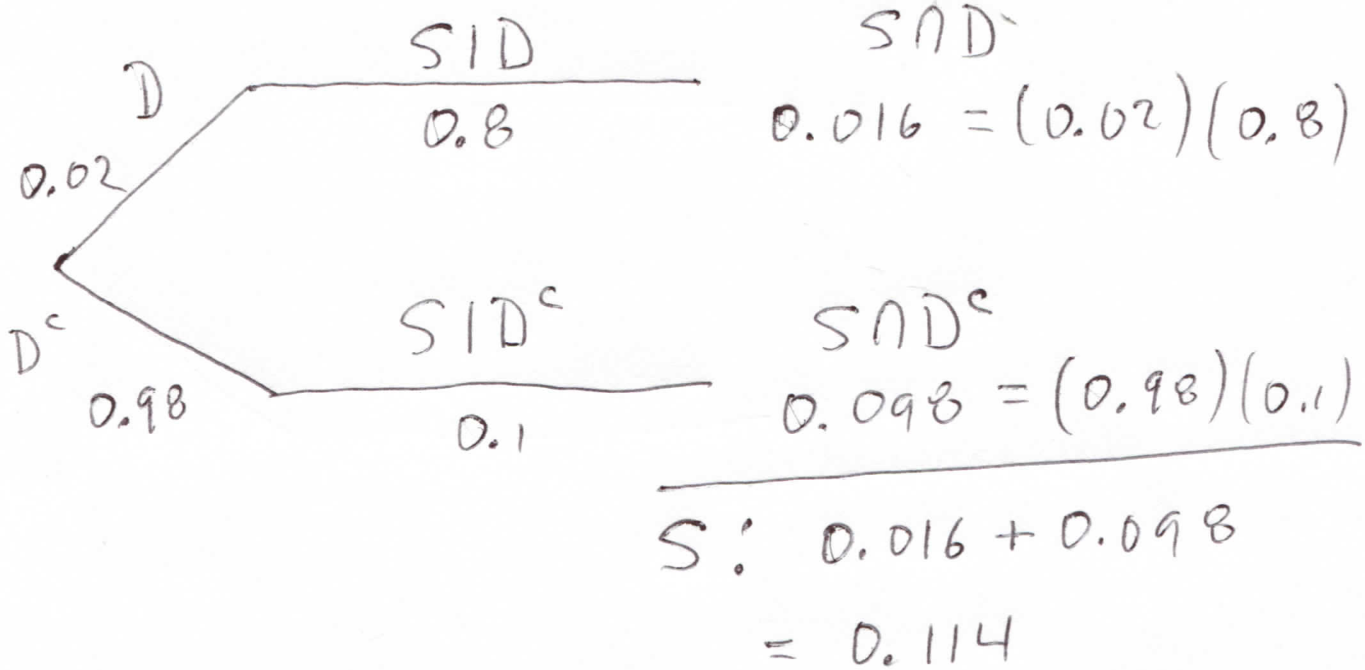
The Multiplication Law 2.7 enables us to fill in probabilities on the right.



Finally, $P(S) = P(S \cap D) + P(S \cap D^c)$ follows from 1.4 or the Addition Law 1.5.



The following filled-in tree diagram has almost any probability relating D and S that could be asked for.



In particular, our desired

$$P(D^c|S) = \frac{P(D^c \cap S)}{P(S)} = \frac{0.098}{0.114} \sim 86\%;$$

an unfortunately high probability that an accusation is false.

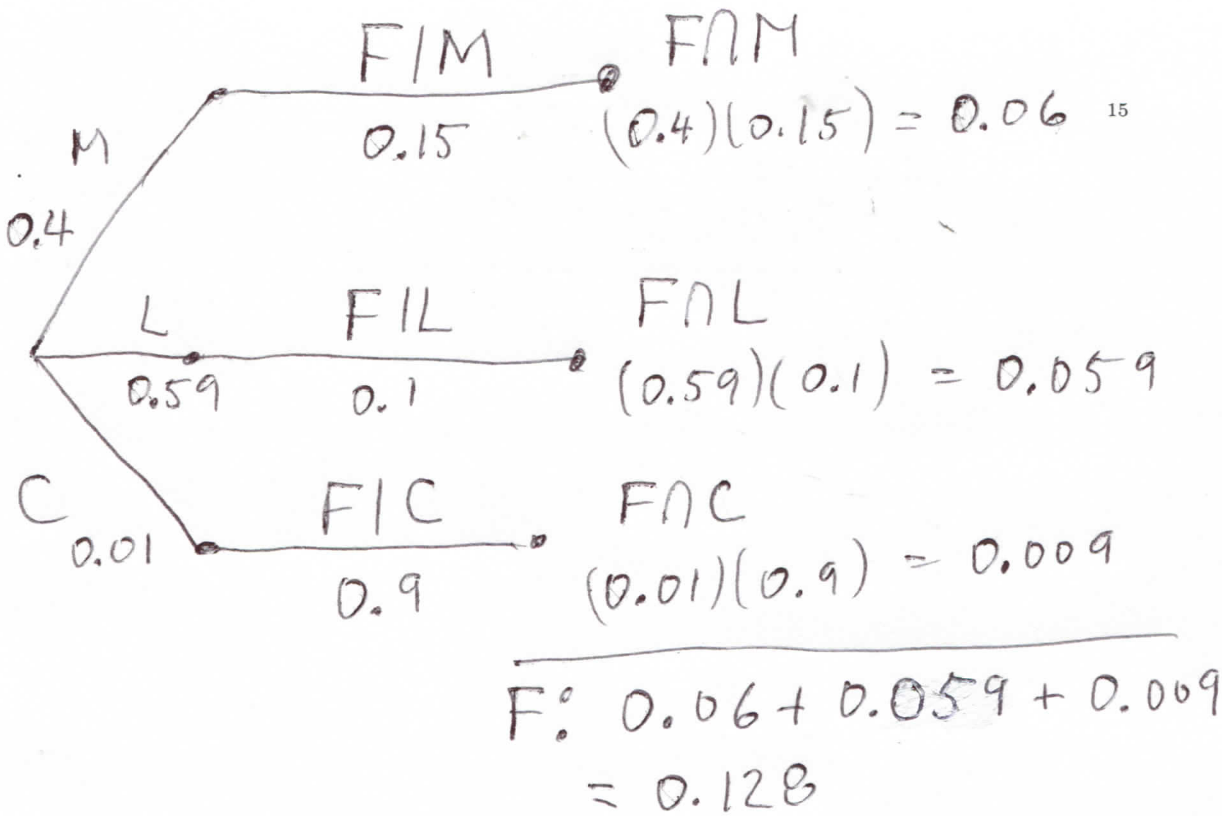
(c) Denote "tent falls" by F, "Moe pitches tent" by M, "Larry pitches tent by" L, "Curly pitches tent" by C.

We have

$$P(M) = 0.4, P(L) = 0.59, P(C) = 0.01, P(F|M) = 0.15, P(F|L) = 0.1, P(F|C) = 0.9,$$

and we want $P(C|F)$.

Put all the given probabilities in a tree, just as in (b), except we have three branches.



We may now read off

$$P(C|F) = \frac{P(C \cap F)}{P(F)} = \frac{0.009}{0.128} \sim 7\%.$$

In other words, only seven percent of tents that fall were pitched by Curly. Accusations of Curly are not looking very fair, even though his track record (ninety percent of tents Curly pitches fall) is terrible.

See [2, Chapter IV] for many more examples of these pictures. In general, we are given $P(A_1), P(A_2), \ldots, P(B|A_1), P(B|A_2), \ldots$, and we want $P(A_1|B), P(A_2|B), \ldots$. In practice, B is something we can see or measure, while $A_1, A_2, \ldots$ are invisible but important; for example, B might be a fever, A_1 a disease, B might be the present while A_1 is the past.

The probabilities that arise from our trees are given names. The **Law of Total Probability** says

$$P(B) = P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + \ldots$$

and **Bayes' Theorem** is

$$P(A_j|B) = \frac{P(B|A_j)P(A_j)}{P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + \ldots},$$

for $j = 1, 2, \ldots$.

We do not recommend using or memorizing either of the formulas just stated; a tree diagram, as in Examples 2.8(b) and (c), is much easier.

REFERENCES

1. Ralph deLaubenfels, "Probability Introduction Magnification," http://teacherscholarinstitute.com/MathMagnificationsReadyToUse.html.

2. Ralph deLaubenfels, "Fun with Introductory Probability," www.teacherscholarinstitute/Books/Probability.pdf