St

Lecture Notes & Examples 5.3

Section 5.3 - Conditional Probability & Independence (pp. 312 - 328)

1. **Conditional** Probability - Let's return to the setting of the homeowners' example in Section 5.2.

	High School Grad	Not a HS Grad	Total
Homeowner	221	119	340
Not a homeowner	89	71	160
Total	310	190	500

If we know that a person owns a home, what is the probability that the person is a high school graduate?

221/340

If we know that a person is a high school graduate, what is the probability that the person owns a home?

221/310

These questions involve **conditional probabilities**. The name comes from the fact that we are trying to find the probability that one event will happen under the *condition* that some other event is already known to have occurred. We often use the phrase ***"given that"*** to signal the condition.

Definition: The probability that one event happens given that another event is already known to have happened is called a **conditional probability**. Suppose we know that event A has happened. Then the **probability that A happens *given that* event B has happened is denoted by P(A | B)**.

Using this notation, we can restate the answers to our two previous questions:

- P(HS grad | Homeowner) = P(HS grad given that they are a Homeowner) 221/340 = $\frac{P(H.S.\ Grad \cap H/O)}{P(H/O)}$
- P(Homeowner | HS grad)= P(Homeowner given that they are a HS grad) 221/310 = $\frac{P(H/O \cap H.S.\ Grad)}{P(H.S.\ Grad)}$

2. Calculating Conditional Probabilities

Conditional Probability Formula

To find the conditional probability P(B | A), use the formula

Prob (B given A)

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

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

Prob A given B

Example: Given the table below, what is the probability that a randomly selected household with a landline also has a cell phone?

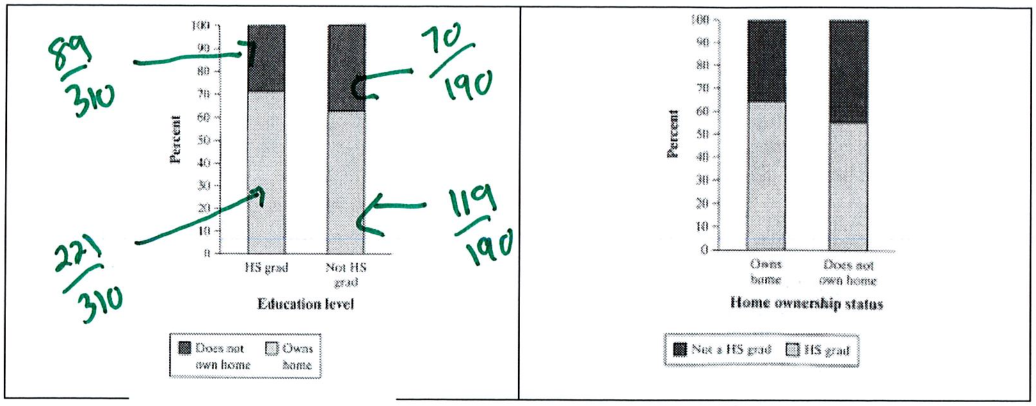
	Cell Phone	No Cell Phone	Total
Landline	0.60	0.18	0.78
No Landline	0.20	0.02	0.22
Total	0.80	0.20	1.00

Given: Choosing from household with LL

$P(CP|LL) = \frac{P(CP \cap LL)}{P(LL)} = \frac{.60}{.78} \approx .769$

Is there a connection between *conditional probability* and the *conditional distribution* from Chapter 1?

The answer is **yes**. The two segmented bar graphs below display the conditional distributions for the Homeowners example.



CHECK YOUR UNDERSTANDING

Students at the University of New Harmony received 10,000 course grades last semester. The two-way table below breaks down these grades by which school of the university taught the course. The schools are Liberal Arts, Engineering and Physical Sciences, and Health and Human Services.

	Grade Level			
School	**A**	**B**	**Below B**	Total
Liberal Arts	2,142	1,890	2,268	6300
Engineering and Physical Sciences	368	432	800	1600
Health and Human Services	882	630	588	2100
Total	3,392	2,952	3656	10000

(This table is based closely on grade distributions at an actual university, simplified a bit for clarity.)

College grades tend to be lower in engineering and the physical sciences (EPS)than in liberal arts and social sciences (which includes Health and Human Services). Consider the two events **E: the grade comes from an EPS course**, and **L: the grade is lower than a B.**

1. Find P(L). Interpret this probability in context.

$P(L) = \frac{3656}{10{,}000} = .3656$

36.56% of grades are lower than a B

2. Find P(E | L) and P(L | E). Which of these conditional probabilities tells you whether this college's EPS students tend to earn lower grades than students in liberal arts and social sciences? Explain.

$P(E|L) = \frac{P(E \cap L)}{P(L)} = \frac{800}{3656} \approx .22$

$P(L|E) = \frac{P(L \cap E)}{P(E)} = \frac{800}{1600} = .50$

P(L|E) gives the prob. of getting a lower grade if the student is in an EPS school.

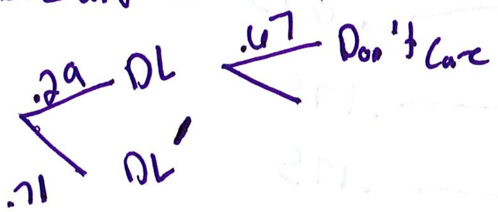
3. The General Multiplication Rule

> **General Multiplication Rule**
>
> The probability that events A and B both occur can be found using the **general multiplication rule**
>
> $$P(A \cap B) = P(A) \cdot P(B|A)$$
>
> Where P(B|A) is the conditional probability that event B occurs given that event A has already occurred.

79. **Free downloads?** Illegal music downloading has become a big problem: 29% of Internet uses download music files, and 67% of downloaders say they don't care if music is copyrighted.[15] What percent of Internet users download music and don't care if its copyrighted? Write the information given in terms of probabilities, and use the general multiplication rule.

P(D.L. and Don't Care) = P(D.L.) · P(Don't Care | DL Music)

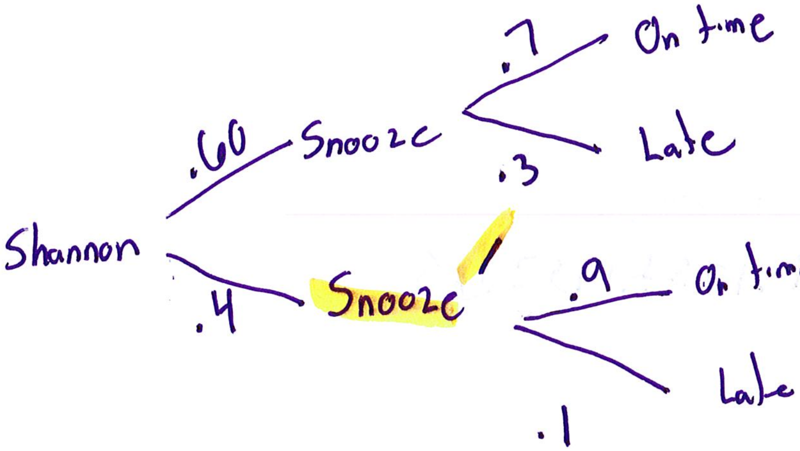
(.29) · (.67) = .1943



4. Tree Diagrams and the General Multiplication Rule

Shannon hits the snooze bar on her alarm clock on 60% of school days. If she does not hit the snooze bar, there is a 0.90 probability that she makes it to class on time. However, if she hits the snooze bar, there is only 0.70 probability that she makes it to class on time. In a randomly chosen day, what is the probability that Shannon is late to class?

P(Late Class)



$P(SN \cap OT) = (.6)(.7) = .42$

$P(SN \cap Late) = .6(.3) = .18$

$P(SN' \cap OT) = .4(.9) = .36$

$P(SN' \cap L) = .4(.1) = .04$

Sum = 1

$P(Late) = P(SN \cap L) + P(SN' \cap L)$

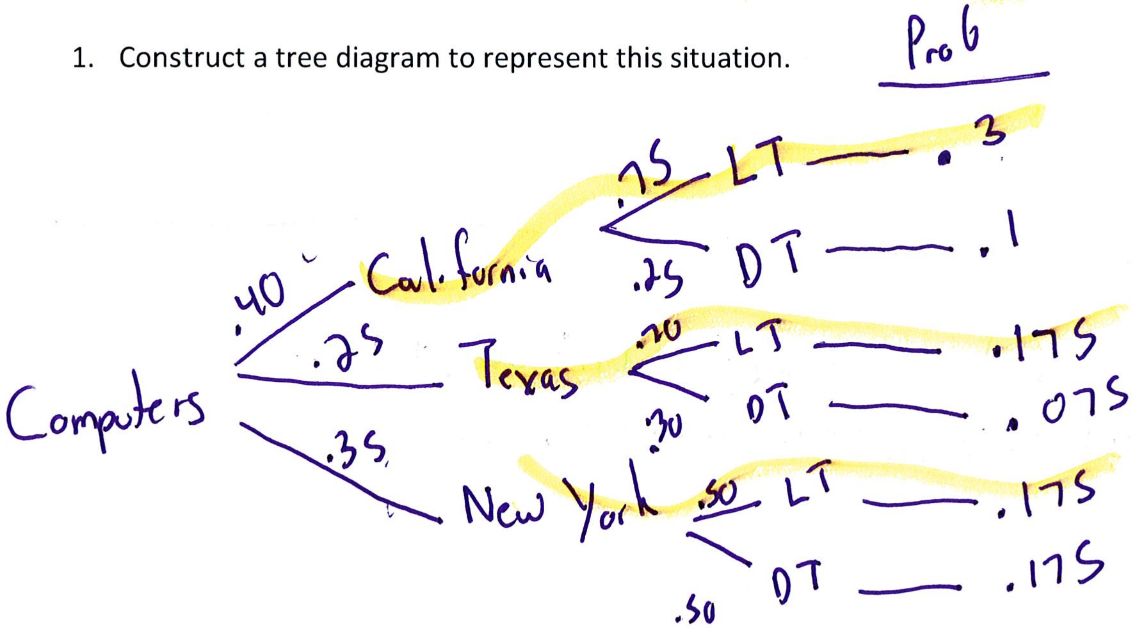
$.18 + .04 = .22$

22% chance of being late.

CHECK YOUR UNDERSTANDING

A computer company makes desktop and laptop computers at factories in three states—California, Texas, and New York. The California factory produces 40% of the company's computers, the Texas factory makes 25%, and the remaining 35% are manufactured in New York. Of the computers made in California, 75% are laptops. Of those made in Texas and New York, 70% and 50%, respectively, are laptops. All computers are first shipped to a distribution center in Missouri before being sent out to stores. Suppose we select a computer at random from the distribution center.

1. Construct a tree diagram to represent this situation.



2. Find the probability that the computer is a laptop. Show your work.

P(laptop) = .30 + .175 + .175 = .65

5. Conditional Probability and Independence

Suppose you toss a fair coin twice. Define events A: first toss is a head, and B: second toss is a head. P(A) = 0.5 and (B) = 0.5. What is P(B|A)? It is the conditional probability that the second toss is a head given that the first toss was a head. The coin has no memory, so P(B|A) = 0.5. In this case P(B|A) = P(B).

Let's contrast the coin-toss scenario with our earlier homeowner example. The events of interest were A: is a high school graduate and B: owns a home. We already learned that P(B) = 340/500 = 0.68 and P(B|A) = 221/310 = 0.712. That is, we know that a randomly selected member of the sample has a 0.68 probability of owning a home. However, if we know that the randomly selected member is a high school graduate, the probability of owning a home increases to 0.712.

Definition. Two events A and B are **independent** if the occurrence of one event has no effect on the chance that the other event will happen. In other words, events A and B are independent if
$P(A|B) = P(A) \text{ and } P(B|A) = P(B)$

Example - Is there a relationship between gender and having allergies? To find out, we used the random the CensusAtSchool web site to randomly select 40 U.S. high school students who completed a survey. The two-way table shows the gender of each student and whether the student has allergies.

	Female	Male	Total
Allergies	10	8	18
No Allergies	13	9	22
Total	23	17	40

Are the events "female" and "allergies" independent?

If independent
$P(A|B) = P(A)$ and $P(B|A) = P(B)$

$P(\text{Allegery} \mid \text{Female}) = \frac{10}{23} \approx .43$

$P(\text{Allergy}) = \frac{18}{40} \approx .45$

"Female" and "Allergies" are not independent

CHECK YOUR UNDERSTANDING

For each chance process below, determine whether the events are independent. Justify your answer.

1. Shuffle a standard deck of cards and turn over the top card. Put it back in the deck, shuffle again, and turn over the top card. Define events a: first card is a heart, and B: second card is a heart.

Independent: 1st card is replaced and this allows 1st event to not affect 2nd event.

2. Shuffle a standard deck of cards, and turn over the top two cards, one at a time. Define events a: first card is a heart, and B: second card is a heart.

Not independent: Once we know the suit of the 1st card, then prob of getting a heart on the 2nd card will change depending on what the 1st card was.

3. The 28 students in Mr. Tabor's AP Statistics class completed a brief survey. One of the questions asked whether each student was right or left-handed. The two-way table summarizes the class data. Choose a student from the class at random. The events of interest are "female" and "right-handed.

	Gender		
Handedness	**Female**	**Male**	
Left	3	1	4
Right	18	6	24
			28

✓ $P(A|B) = P(A)$

$P(\text{R.H} \mid \text{Female}) = \frac{18}{21} = \frac{6}{7}$

$P(\text{R.H}) = \frac{24}{28} = \frac{6}{7}$

$\therefore$ Independent

5. Independence: A Special Multiplication Rule - What happens to the general multiplication rule when events A and B are independent?

Multiplication Rule for Independent Events
If A and B are independent events, then the probability that A and B both occur is
$P(A \cap B) = P(A) \cdot P(B)$

Example: In baseball, a perfect game is when a pitcher does not allow any hitters to reach base in all nine innings. Historically, pitchers throw a perfect inning - an inning where no hitters reach base - about 40% of the time. So, to throw a perfect game, a pitcher needs to have nine perfect innings in a row.

What is the probability that a pitcher throws nine perfect innings in a row, assuming the pitcher's performance in an inning is independent of his performance in the other innings.

$P(9 \text{ perfect innings in a row}) = \underset{1}{.40} \cdot \underset{2}{.40} \cdot \underset{3}{\quad} \cdots \cdot \underset{9}{.40}$

$= (.40)^9 = .00026$

Example: The First Trimester Screening is a noninvasive test given during the first trimester of pregnancy to determine if there are specific chromosomal abnormalities in the fetus. According to the *New England Journal of Medicine* in November 2005, approximately 5% of normal pregnancies will receive a false positive result.

Reasonable to assume that test results of different women independent

Among 100 women with normal pregnancies, what is the probability that there will be *at least* 1 false positive?

P(at least 1 false positive) = 1 - P(no false pos)

for a single test

P(false pos) = .05   Prob(not false pos) = 1 - .05 = .95

$P(100 \text{ not false pos}) = \underset{1}{(.95)}\underset{2}{(.95)} \ldots \underset{100}{(.95)}$ (100 times) $= (.95)^{100} = .0059$

prob all 100 women will get not false pos

Prob(at least 1 false pos for 100 women) = 1 - .0059 = .9941

about 99.41% chance of at least 1 false positive for every 100 women.

CHECK YOUR UNDERSTANDING

1. During World War II, the British found that the probability that a bomber is lost through enemy action on a mission over occupied Europe was 0.05. Assuming that missions are independent, find the probability that a bomber returned safely from 20 missions.

P(bomber lost) = .05

P(return safely) = .95

P(return safely from 20 missions) = $.95^{20}$

= .3585

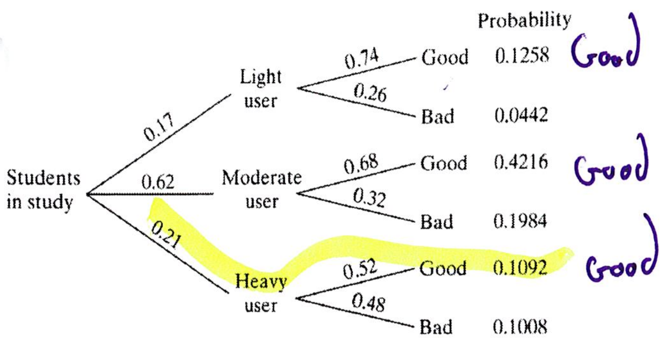
2. Government data show that 8% of adults are full-time college students and that 30% of adults are age 55 or older. Since (0.08)(0.30) = 0.024, can we conclude that about 2.4% of adults are college students 55 or older? Why or why not?

No whether one is a college student and one's age are not independent.

Far more younger people than older people are college students.

Given

Example: Given the diagram below, what percent of youth with good grades are heavy users of media?



P(heavy users | Good Grades)

$$= \frac{P(h.u \cap G.G)}{P(GG)} = \frac{.1092}{.1258 + .4216 + .1092}$$

$$= \frac{.1092}{.6566} = .166$$

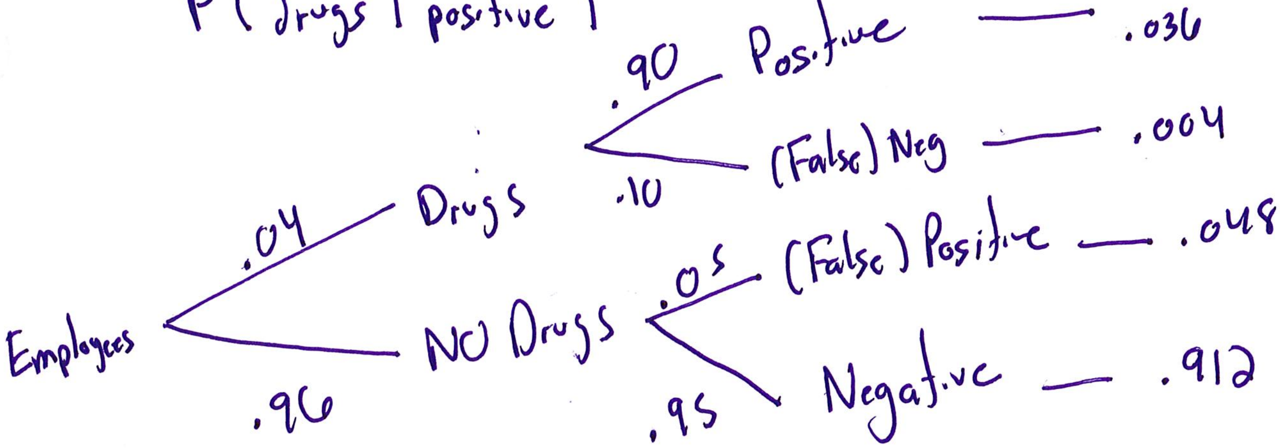
16.6 % of youth with good grades are heavy users of social media.

Example: Many employers require prospective employees to take a drug test. A positive result indicates that the prospective employee uses illegal drugs. However, not all people who test positive actually use drugs. Suppose that 4% of the prospective employees use drugs, the false positive rate is 5% and the false negative rate is 10%.

What percent of people who test positive actually use drugs?

P(drugs | positive)

P(Drugs) = .04



$$= \frac{P(D \cap Pos)}{P(Pos)} = \frac{.036}{.036 + .048} = \frac{.036}{.084} \approx .43$$