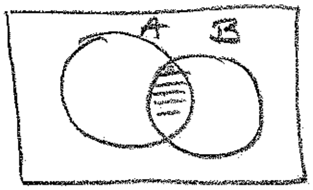
Probability Review

- The probability of event A is denoted: $P(A)$
- disjoint events or mutually exclusive events means that the events cannot happen at the same time or in sequence consecutively.
- The General Addition Rule:

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



Intersection counted twice.
Subtract it out once.

- Independent events means that if one event occurs, it does not change the probability that the other event occurs.
- The General Multiplication Rule:

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

- A Conditional probability is the probability that event B occurs if event A has already occurred.
- We denote conditional probability as: $P(B|A)$.
- Two events are independent if:

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

Event A does not matter.

- Probabilities can be given in contingency tables, Venn Diagrams, in words, or estimated from raw data (empirical method), using equally likely outcomes (classical method), or based on personal beliefs and experiences (subjective method).

Example: Six people aspire to be chosen to be on a game show. Only two will be selected, and the producers will select the contestants completely at random. Two of the people are younger than 25 years old.

a. What is the sample space of all possible outcomes for this probability experiment?

Call the contestants A, B, C, D, E, F

AB	BC	CD	DE	EF
AC	BD	CE	DF	
AD	BE	CF		
AE	BF			
AF				

b. What is the probability that both contestants are younger than 25? Compute in two ways.

Let contestant A, B be under 25.

Then above, P(Both under 25) = $\frac{1}{15}$

OR P(Both Under 25) = P(1st under 25) × P(2nd under 25 given 1st is)

$= \left(\frac{2}{6}\right) \times \left(\frac{1}{5}\right) = \frac{2}{30} = \frac{1}{15}$

c. What is the probability that one is younger than 25 and the other is not? Compute in two ways.

All combos with A or B but not both

$\frac{8}{15}$

Reverse order

OR P(one under and one over) =

P(1st under) × P(2nd over | 1st under) × 2 =

$= \frac{2}{6} \times \frac{4}{5} \times 2 = \frac{16}{30} = \frac{8}{15}$

- The complement of an event is the set of all outcomes not in the event.
- The complement rule: $P(A^c) = 1 - P(A)$

Example: Carlito drives through two traffic lights on his way to work. Denote E = "Gets stopped at the first light" with probability of 0.40 and denote F = "Gets stopped at the second light" with probability 0.30. The $P(E \text{ and } F) = 0.15$.

a. Set up a two-by-two contingency table to display all possible outcomes.

	E	E^c	TOTAL
F	0.15	0.15	0.30
F^c	0.25	0.45	0.70
	0.40	0.60	1.00

b. What is the probability that Carlito must stop for at least one light?

$$P(E \text{ OR } F) = P(E) + P(F) - P(E \text{ AND } F)$$
$$= 0.40 + 0.30 - 0.15 = 0.55$$

c. What is the probability he doesn't have to stop at either light?

$$P(E^c \text{ and } F^c) = 0.45$$

d. What is the probability he stops at exactly one light?

$$P(E \text{ AND } F^c) + P(E^c \text{ AND } F) = 0.25 + 0.15$$
$$= 0.40$$

e. If Carlito works a condensed four-day week, what is the probability he goes four days without getting stopped even once?

$P(E^c \text{ AND } F^c) = 0.45$ for any day

$P(4 \text{ days in a row no stops}) =$

$$0.45 \times 0.45 \times 0.45 \times 0.45 =$$

$$(0.45)^4 = 0.041$$

Example: Suppose events A and B are mutually exclusive with $P(A) = 0.41$ and $P(B) = 0.23$.

a. What is $P(A \text{ and } B)$? $P(A \text{ and } B) = 0$

b. What is $P(A \text{ or } B)$? $P(A \text{ OR } B) = P(A) + P(B)$

$= 0.41 + 0.23 = 0.64$

Conditional Probability Formula:

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

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

Example: We know at Cecil College that about 65% of our students are female and that 90% of the students in our Nursing Program are female. For the whole population of students at Cecil, about 4% are in the nursing program. ~~Let A = "Student is female" and B = "Student is in the Nursing~~ Program".

a. What is the probability that a randomly selected Cecil student is a female in the Nursing Program?

$P(\text{Female}) = 0.65 \quad P(\text{Nursing}) = 0.04$

$P(\text{Female} | \text{Nursing}) = 0.90$

$P(\text{Female AND Nursing}) = P(\text{Female} | \text{Nursing}) P(\text{Nursing})$

$= (0.90)(0.04) = 0.036$

b. What is the probability that three randomly selected students are all female?

$P(\text{all 3 female}) = P(1^{st} \text{ and } 2^{nd} \text{ and } 3^{rd} \text{ F})$

$= (0.65)^3 = 0.2746$

Treat as Indep.

c. What is the probability that in a group of 5 randomly selected students, at least one of them is in the Nursing Program?

P(at least one in Nursing) =

1 − P(none are in nursing) =

1 − P(not AND NOT AND ....) =

$1-(0.96)^5 = 0.1846$

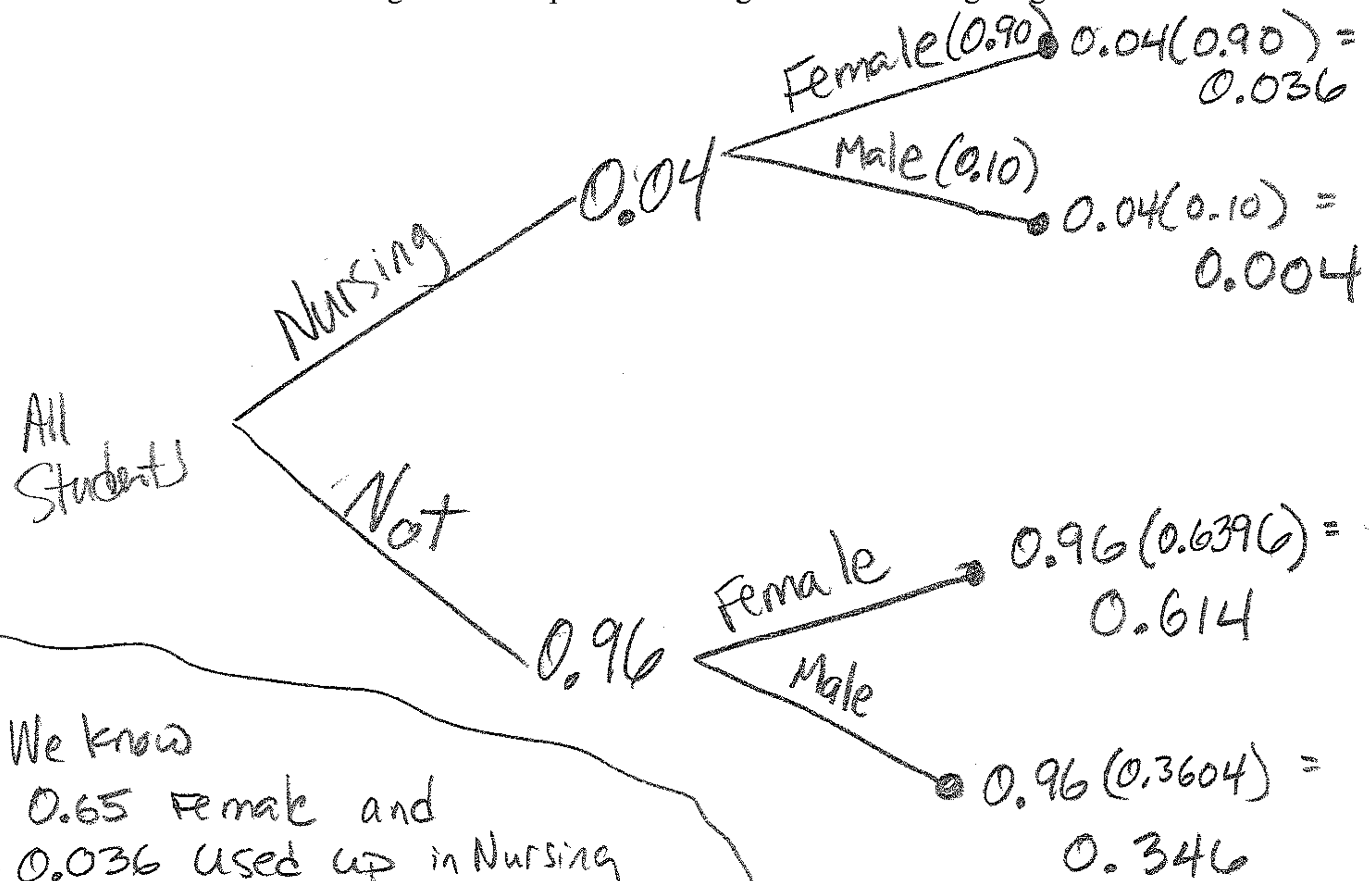
18.46% chance at least one in Nursing

d. Are gender and being in the Nursing Program independent or dependent? Why?

Dependent -

P(Female) = 0.65 ≠ P(Female | Nursing) = 0.90

e. Draw a tree diagram for the possibilities of gender and Nursing Program.



We know
0.65 Female and
− 0.036 used up in Nursing

0.614

f. Determine the probability that a student is in nursing if we know she is female.

$$P(\text{Nursing} \mid \text{Female}) = \frac{P(\text{Nursing And Female})}{P(\text{Female})}$$

$$= \frac{0.036}{0.65} = 0.0554$$

5.54%

g. Determine the probability that a student is in nursing if we know he is male.

$$P(\text{Nursing} \mid \text{Male}) = \frac{P(\text{Nursing AND Male})}{P(\text{Male})}$$

$$= \frac{0.004}{0.35} = 0.0114$$

1.14%

Bayes' Rule

We have $P(A|B)$ but want $P(B|A)$.

$$P(B|A) = \frac{P(A|B)P(B)}{P(A|B)P(B) + P(A|B^c)P(B^c)}$$

← Both A and B

↖ P(A)

Example: Lyme disease is the leading tick-borne disease in the US. Diagnosis of the disease is difficult and is aided by a test that detects certain antibodies in the blood.

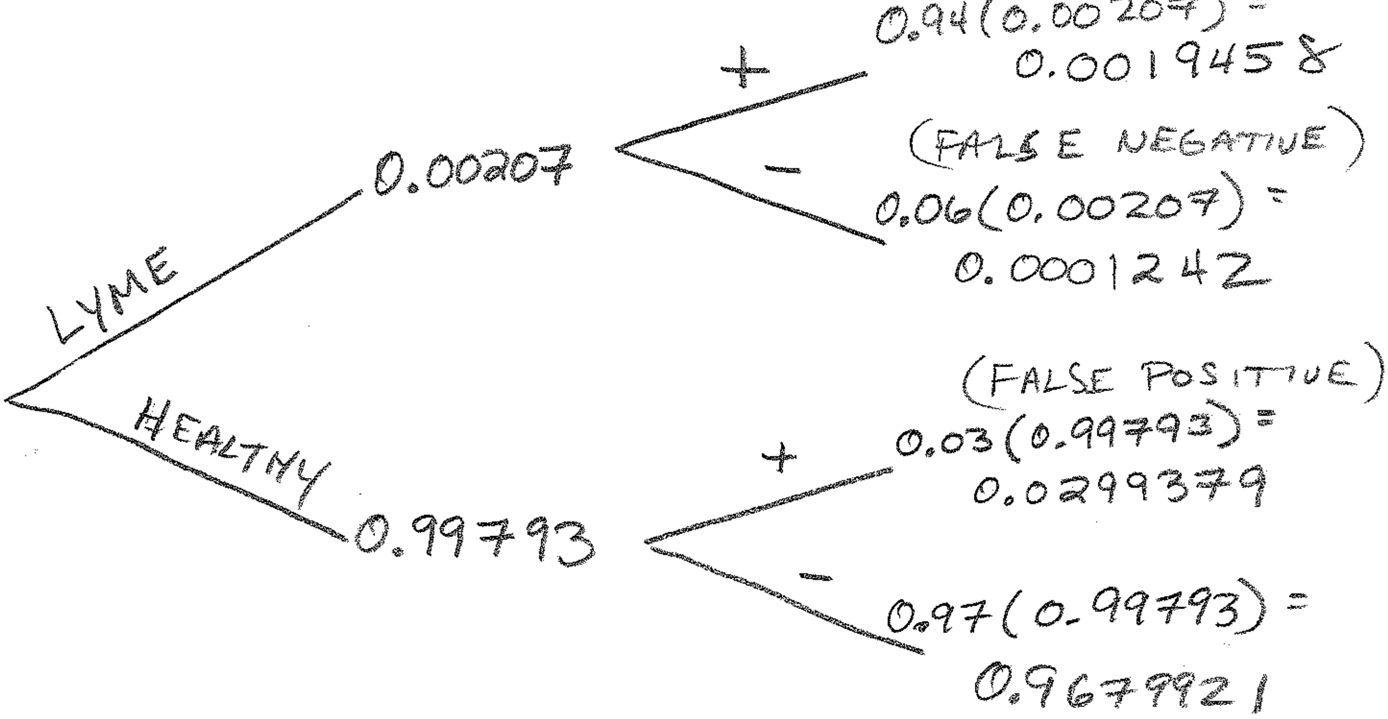
Let + represent a positive result on the blood test and – represent a negative result on a blood test.

L represents that a patient actually has Lyme disease and L^C means that they don't.

A journal article reports the following probabilities:

$P(L) = 0.00207$ and $P(L^C) = 0.99793$ with a false positive rate of 3% and a false positive rate of 6%.

Create a tree diagram and then determine the probability that a person actually has Lyme disease if they test positive.



P(HAS LYME | TEST POSITIVE) =

$$\frac{P(\text{HAVE LYME AND TEST POSITIVE})}{P(\text{TEST POSITIVE})} =$$

$$\frac{0.0019458}{0.0019458 + 0.0299379} = 0.061$$

6.1% chance

Example: Let us assume 20% of adults smoke cigarettes. We know that 60% of smokers and 10% of nonsmokers will develop a lung condition by age 55.

$P(\text{Smoke}) = 0.2$
$P(\text{Don't}) = 0.8$

a. Explain how these figures support that lung conditions and smoking are not independent.

b. What is the probability that a 55-year-old person has a lung condition?

c. What is the probability ~~that someone~~ with a lung condition was a smoker?

a.) Very simply $P(\text{Lung Condition} \mid \text{smoke}) = 0.60$ is not equal to $P(\text{Lung Condition} \mid \text{Don't}) = 0.10$.

b.) $P(\text{55 y/o has lung condition}) =$

$$P(\text{55+ has condition} \mid \text{Smoke})P(\text{smoke}) +$$
$$P(\text{55+ has condition} \mid \text{Don't})P(\text{Don't}) =$$

$$(0.60)(0.20) + (0.10)(0.80) = 0.20 = 20\%$$

c.) $P(\overset{B}{\text{Smoker}} \mid \overset{A}{\text{Lung}}) =$ Bayes

$$\frac{P(\overset{A}{\text{Lung Condition}} \mid \overset{B}{\text{Smoke}})P(\overset{B}{\text{smoke}})}{P(\text{Lung Condition} \mid \text{Smoke})P(\text{Smoke}) + P(\text{Lung} \mid \text{Don't})P(\text{Don't})} =$$

$$\frac{(0.60)(0.20)}{(0.60)(0.20) + (0.10)(0.80)} =$$

$$\frac{0.12}{0.20} = 0.60$$

60% chance was a smoker.

Example: HIV testing is not foolproof, nor is it 100% reliable. At a clinic for HIV testing, we know that 15% of the patients have HIV. We know from large-scale testing that 99.7% of people with HIV test positive. Also, 98.5% of people without HIV test negative.

a. Write out in symbols as many probabilities as we can from the above information.
b. What is the probability that a person who tests negative is actually free of HIV?
c.) Added.

a.) $P(HIV) = 0.15$ $\quad P(+ \mid HIV) = 0.997$
$P(Healthy) = 0.85$ $\quad P(- \mid HIV) = 0.003$ (False neg)

$P(+ \mid Healthy) = 0.015$ (False pos)
$P(- \mid Healthy) = 0.985$

b.) $P(Healthy \mid -) =$ (Bayes)

$$= \frac{P(- \mid Healthy)\,P(Healthy)}{P(- \mid Healthy)\,P(Healthy) + P(- \mid HIV)\,P(HIV)}$$

$$= \frac{(0.985)(0.85)}{(0.985)(0.85) + (0.003)(0.15)}$$

$$= \frac{0.83725}{0.8377} = 0.9995$$

99.95% chance of being healthy if you test negative.

c.) $$P(HIV \mid +) = \frac{P(+ \mid HIV)\,P(HIV)}{P(+ \mid HIV)\,P(HIV) + P(+ \mid Healthy)\,P(Healthy)}$$

$$= \frac{(0.997)(0.15)}{(0.997)(0.15) + (0.015)(0.85)} = 0.921$$

92.1%

Two Bayes' Stories (From Nate Silver's book *The Signal and the Noise*)

- Suppose you are living with a partner and come home from a business trip to find a strange pair of underwear in your dresser drawer. For this example, you are the woman and your husband was the one at home while you were away.
- The natural question is: "Was your husband cheating on you?"
- The condition is that you found the underwear.
- The hypothesis you are interested in testing is the probability that you are being cheated on.
- Bayes' theorem can answer this sort of question if we are willing to estimate three quantities.
- We need to estimate the probability of the underwear's appearing, as a condition of the hypothesis being true. That is, you are being cheated upon. If he's cheating on you, you'd expect him to be extra careful. If he's cheating on you, it's easy enough to imagine how they got there in the first place.
  - Let's put the probability of the underwear appearing there, conditional on his cheating on you, at 50%.
- We need to estimate the probability of the underwear's appearing, Conditional on the hypothesis being false. If he isn't cheating, are there some innocent explanations for how they got there? They could be his underwear. It could be luggage got mixed up. It could be a gift to you. It could be from a platonic friend who you trust. Very unlikely mix (like the dog eating your homework), but not impossible.
  - Let's put the probability of the underwear appearing there, conditional on his not cheating on you, at 5%.
- Bayesians need a *prior probability*. What is the probability that you would have assigned to him cheating on you before you found the underwear (it would be hard to be objective at this point). Studies have found that 4 % of married partners cheat on their spouses in any given year, so set that as our prior.

- Now use Bayes' formula to determine the probability of him cheating on you, *given* that you found the underwear:

$$P(\text{Cheating} \mid \text{Found Underwear}) =$$

$$\frac{P(\text{Found Underwear} \mid \text{Cheat})\,P(\text{Cheat})}{P(\text{Found} \mid \text{Cheat})P(\text{Cheat}) + P(\text{Found} \mid \text{Not})P(\text{Not})} =$$

$$\frac{(0.50)(0.04)}{(0.50)(0.04) + (0.05)(0.96)} =$$

$$0.2941 \quad \text{or} \quad 29.41\%$$

- This may seem counterintuitive, but it stems from the fact that we assigned a very low prior probability to him cheating. We started out thinking he was an __innocent__ man, so that weighs heavily into our equation.

- Consider a second example – the September 11 attacks. Most everyone put a very low prior probability to terrorists crashing planes into buildings when we woke up that morning.

- Say, before the first plane hit, we put the estimate of a terror attack on tall buildings in Manhattan at just __1 in 20,000__, or __0.00005__.

- Accidents do happen, and in the previous 25,000 days in Manhattan, two accidents like these did occur (Empire State Building 1945 and 40 Wall Street in 1946). The probability of an accident would then be about __1 out of 12,500__ on any given day, based on historical empirical evidence.

OR 0.00008

- The new event unfortunately occurred: The first plane hits The World Trade Center. Use Bayes' Theorem to calculate the probability that it was a terrorist attack:

$$P(\text{attack} \mid \text{hit}) = \frac{P(\text{hit} \mid \text{attack})P(\text{attack})}{P(\text{hit} \mid \text{attack})P(\text{attack}) + P(\text{hit} \mid \text{accident})P(\text{accident})} = \frac{(1)(0.00005)}{(1)(0.00005) + (0.00008)(0.99995)}$$

$$\approx 0.38 \text{ (about 38\% chance its an attack)}$$

- Bayes' allows us to continually update our probability estimates as new evidence presents itself. Our __posterior__ probability of a terror attack after the first plane hit, 38%, becomes our new __prior__ probability before the second plane did.

- Re-compute Bayes' to get the probability we were under attack after the second plane hit. We learn that one accident was unlikely enough, but a second one was almost a literal impossibility:

$$P(\text{attack} \mid \text{2 hits}) = \frac{P(\text{2 hits} \mid \text{attack})P(\text{attack})}{P(\text{2 hits} \mid \text{attack})P(\text{attack}) + P(\text{2 hits} \mid \text{accident})P(\text{accident})}$$

$$= \frac{(1)(0.38)}{(1)(0.38) + (0.00008)^2(0.62)}$$

$$= 0.999999896 \text{ or near certainty.}$$