Stat 250 Gunderson Lecture Notes
3: Probability

Chance favors prepared minds. -- *Louis Pasteur*

Many decisions that we make involve uncertainty and the evaluation of probabilities.

Interpretations of Probability

Example: Roll a fair die → possible outcomes = { }

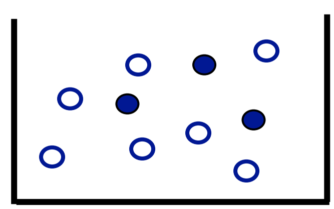
Before you roll the die do you know which one will occur?
What is the probability that the outcome will be a '4'? __________ Why?

A few ways to think about PROBABILITY:

(1) Personal or Subjective Probability
P(A) = the degree to which a given individual believes that the event A will happen.

(2) Long term relative frequency
P(A) = proportion of times 'A' occurs if the random experiment (circumstance) is repeated many, many times.

(3) Basket Model
P(A) = proportion of balls in the basket that have an 'A' on them.



10 balls in the basket: 3 blue and 7 white
One ball will be selected at random.

What is P(blue)? ______________

Note: A probability statement ***IS NOT*** a statement about __________________ .

It ***IS*** a statement about __________________________ .

Discover Basic Rules for Finding Probability through an Example

There is a lot you can learn about probability. One basic rule to always keep in mind is that the probability of any outcome is always between 0 and 1. Now, there are entire courses devoted just to studying probability. But this is a Statistics class. So rather than start with a list of definitions and formulas for finding probabilities, let's just do it through an example so you can see what ideas about probability we need to know for doing statistics.

Example: Shopping Online

Many Internet users shop online. Consider a population of 1000 customers that shopped online at a particular website during the past holiday season and their results regarding whether or not they were satisfied with the experience and whether or not they received the products on time. These results are summarized below in table form. Using the idea of probability as a proportion, try answering the following questions.

	On Time	**Not On Time**	***Total***
Satisfied	800	20	820
Not Satisfied	80	100	180
Total	880	120	1000

a. What is the probability that a randomly selected customer was satisfied with the experience?

b. What is the probability that a randomly selected customer was *not* satisfied with the experience?

c. What is the probability that a randomly selected customer was both satisfied *and* received the product on time?

d. What is the probability that a randomly selected customer was either satisfied *or* received the product on time?

e. *Given* that a customer did receive the product on time, what is the probability that the customer was satisfied with the experience?

f. *Given* that a customer did *not* receive the product on time, what is the probability that the customer was satisfied with the experience?

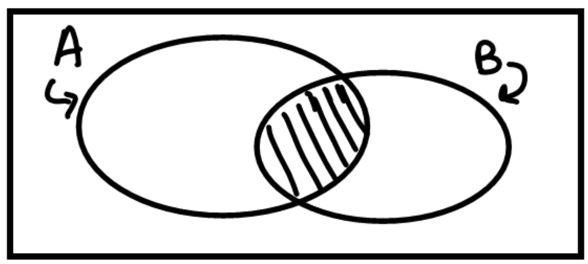
Note*: We stated the above 1000 customers represented a population. If results were based on a sample that is representative of a larger population, then the observed sample proportions would be used as approximate probabilities for a randomly selected person from the larger population.*

Great job! You just computed probabilities using many of the basic probability rules or formulas summarized below and also found in your textbook.

- **Complement rule** $P(A^C) = 1 - P(A)$
- **Addition rule** $P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$
- **Multiplication rule** $P(A \text{ and } B) = P(A)P(B \mid A)$
- **Conditional Probability** $P(A \mid B) = \dfrac{P(A \text{ and } B)}{P(B)}$

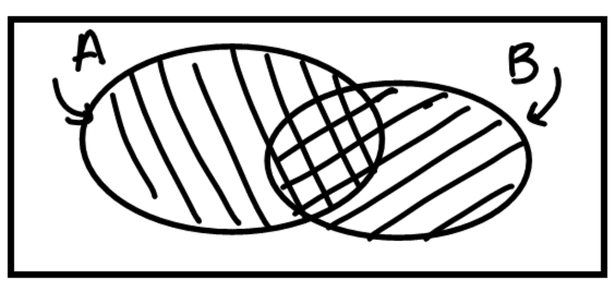
You did not need the formulas themselves but instead used intuition and approaching it as some type of proportion. Let's see how your intuition and the above formulas really do connect.

In part b you found the probability of "NOT being satisfied", which is the complement of the event "being satisfied", so the answer to part b is the complement of the probability you found in part a.



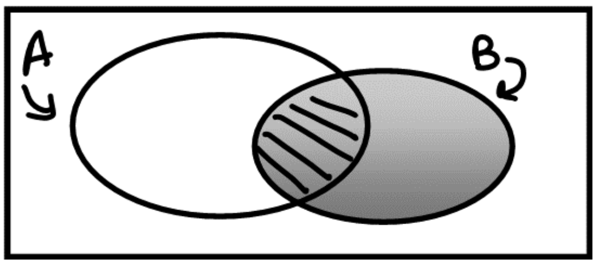
In part c, there was a key word of "AND" in the question being asked. The "AND" is just the intersection, or the overlapping part; the outcomes that are in common. The picture at the right show an intersection between the event A and B. In a table, the counts that are in the middle are the "AND" counts; there were 800 (out of the 1000 customers) that were *both* satisfied AND received the product on time. There is a *multiplication* formula above for finding probabilities of the AND or intersection of two events, but we did not even need to apply it; as a table presentation of counts provides "AND" counts directly.

In part d, there was a key word of "OR" in the question being asked. The "OR" is union, the outcomes that are in either one or the other (including those that are in both). The picture at the right show an union between the event A and B. Notice that if you start with all of the outcomes that are in A and then add all of the outcomes that are in B, you have double counted the outcomes that are in the overlap. So the *addition* formula above shows you need to subtract off the intersecting probability once to correct for the double counting. From the table, you could either add up the separate counts of 800 + 20 + 80; or start with the 820 that were satisfied and add the 880 that received it on time and then subtract the 800 that were in both sets; to get the 900 in all that were either satisfied *or* received the product



on time.

Finally, parts e and f were both conditional probabilities. In part e you were first told to consider only the 880 customers that received the product on time, and out of these find the probability (or proportion) that were satisfied. There were 800 out of the 880 that were satisfied. The picture below shows the idea of a conditional probability formula above for P(A | B), read as the probability of A given B has occurred. If we know B has occurred, then only look at those items in the event B. The event B, shaded at the right, is our new 'base' (and thus is in the denominator of the formula). Now out of those items in B, we want to find the probability of A. The only items in A that are on the set B are those in the overlap or intersection. So the *conditional* formula above shows you count up those in the "A and B" and divide it by the base of "B".

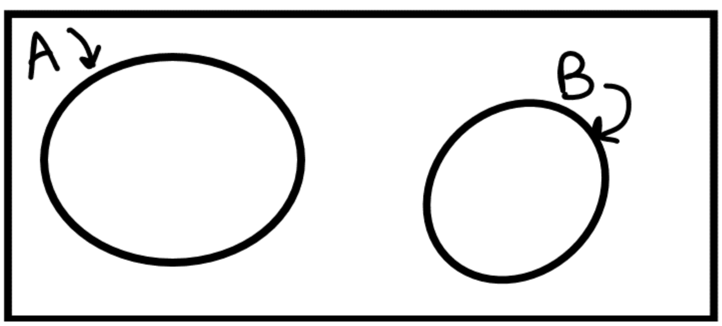


Try It! Go back to parts a to f and add the corresponding shorthand probability notation of what you actually found; e.g. **P(satisfied), P(satisfied | on time)** next to each answer.

Now there are a couple of useful situations that can make computing probabilities easier.

> ***Definition:***
> Two events A, B are ***Mutually Exclusive*** (or ***Disjoint***) if ... **they do not contain any of the same outcomes. So their intersection is empty.**

We can easily *picture* disjoint events because the definition is a property about the sets themselves.



If A, B are disjoint, then P(A and B) = 0. If there are no items in the overlapping part, then man of the probability results will simplify. For example, the additional rule for disjoint events P(A or B) = P(A) + P(B).

Another important situation in statistics occurs when the two events turn out to be ***<u>independent</u>***.

> ***Definition:***
> Two events A, B are said to be **independent** if knowing that one will occur (or has occurred) does not change the probability that the other occurs.
> In notation this can be expressed as **P(A|B) = P(A).**

This expression P(A|B) = P(A) tells us that knowing the event B occurred does not change the probability of the event A happening.

Now it works the other way around too, if A and B are independent events, then P(B|A) = P(B).

As a result of this independence definition, we could show that the <u>multiplication rule for independent events</u> reduces to P(A and B) = P(A)P(B).

Finally, this rule can also be extended. If three events A, B, C are all independent then P(A and B and C) = P(A)P(B)P(C).

So let's apply these two new concepts to our online shopping example.

Back to the Shopping Online Example

Below are results for a population of 1000 customers that shopped online at a particular website during the past holiday season. Recorded was whether or not they were satisfied with the experience and whether or not they received the products on time.

	On Time	**Not On Time**	***Total***
Satisfied	800	20	820
Not Satisfied	80	100	180
Total	880	120	1000

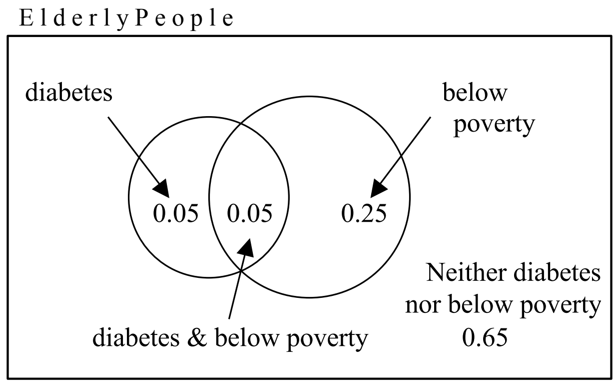
g. Are being satisfied with the experience and receiving the product on time *mutually exclusive* (*disjoint*)? Provide support for your answer.

h. Are being satisfied with the experience and receiving the product on time statistically *independent*? Provide support for your answer.
Hint: go back and compare your answers to parts a, e, and f.

Try It! Elderly People

Suppose that in a certain country, 10% of the elderly people have diabetes. It is also known that 30% of the elderly people are living below poverty level and 5% of the elderly population falls into both of these categories.

At the right is a diagram for these events. Do the probabilities make sense to you?



a. What is the probability that a randomly selected elderly person is not diabetic?

b. What is the probability that a randomly selected elderly person is either diabetic *or* living below poverty level?

c. Given a randomly selected elderly person is living below poverty level, what is the probability that he or she has diabetes?

d. Since knowing an elderly person lives below the poverty level (circle one)

DOES **DOES NOT** change the probability that they are diabetic, the two

events of living below the poverty level and being diabetic (circle one)

ARE **ARE NOT** independent.

In the next example, you are not asked to determine if two events are independent, but rather put independence to use.

Try It! Blood Type
About 1/3 of all adults in the United States have type O+ blood. Suppose three adults will be randomly selected.
Hint: randomly selected implies the results should be________________________.

What is the probability that the **first** selected adult will have type O+ blood?

What is the probability that the **second** selected adult will have type O+ blood?

What is the probability that **all three** will have type O+ blood?

What is the probability that **none** of the three will have type O+ blood?

What is the probability that **at least one** will have type O+ blood?

Some final notes…

I. Sampling with and without Replacement

> ***Definitions:***
> A sample is drawn **with replacement** if individuals are returned to the eligible pool for each selection. A sample is drawn **without replacement** if sampled individuals are not eligible for subsequent selection.

If sampling is done with replacement, the Extension of Rule 3b holds. If sampling is done without replacement, probability calculations can be more complicated because the probabilities of possible outcomes at any specific time in the sequence are conditional on previous outcomes.

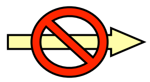
If a sample is drawn from a very large population, the distinction between sampling with and without replacement becomes unimportant. In most polls, individuals are drawn without replacement, but the analysis of the results is done as if they were drawn with replacement. The consequences of making this simplifying assumption are negligible.

II. Sometimes students confuse the mutually exclusive with independence.

Check the definitions.

- The definition for two events to be ***disjoint*** (mutually exclusive) was based on a **SET** property.
- The definition for two events to be ***independent*** is based on a **PROBABILITY** property.

You need to check if these definitions hold when asked to assess if two events are disjoint, or if two events are independent.

Mutually Exclusive  **Independence**

If two events are mutually exclusive then we know that P(A and B) = 0.
This also implies that P(A|B) is equal to 0 (if the two events are disjoint and B did occur, then the chance of A occurring is 0).

So P(A|B) (which is 0) will not be equal to P(A) if the events are disjoint.

III. Probability rules summary

Below is a summary of the key probability results you need to understand and be able to use.

- **Complement rule** $P(A^C) = 1 - P(A)$
- **Mutually Exclusive (disjoint) Events:**
  The events A, B are disjoint if "A and B" is the empty set.
  Thus, P(A and B) = 0.
- **Addition Rule (general)** $P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$
  If A, B are disjoint, we have $P(A \text{ or } B) = P(A) + P(B)$
- **Conditional Probability (general)** $P(A \mid B) = \dfrac{P(A \text{ and } B)}{P(B)}$
- **Independent Events:**
  The events A, B are independent if $P(A \mid B) = P(A)$
  Equivalently, the events A, B are independent
  if $P(A \text{ and } B) = P(A)P(B)$

The Stats 250 formula card provides a more extensive list, but remember, you may not need them as you discovered in your first probability example with the online customers.

Additional Notes

A place to … jot down questions you may have and ask during office hours, take a few extra notes, write out an extra problem or summary completed in lecture, create your own summary about these concepts.