14.31/14.310 Lecture 2

Probability

Let's start out with some definitions

A sample space S is a collection of all possible outcomes of an experiment.

An event A is any collection of outcomes (including individual outcomes, the entire sample space, the null set).

If the outcome is a member of an event, the event is said to have occurred.

Event B is contained in event A is every outcome in B also belongs to A.

Probability

These are sets, so we use many of the same definitions (e.g., intersection, union, complement), and all results from set theory apply (e.g., associative, commutative, distributive properties). Here are some other useful results:

If $A \subset B$ then $A \cup B = B$

If $A \subset B$ and $B \subset A$ then $A = B$

If $A \subset B$ then $AB = A$

$A \cup A^c = S$

Probability

These are sets, so we use many of the same definitions (e.g., intersection, union, complement), and all results from set theory apply (e.g., associative, commutative, distributive properties). Here are some other useful results:

This is how we indicate "contained in"

If $A \subset B$ then $A \cup B = B$

If $A \subset B$ and $B \subset A$ then $A = B$

If $A \subset B$ then $AB = A$

$A \cup A^c = S$

Probability

These are sets, so we use many of the same definitions (e.g., intersection, union, complement), and all results from set theory apply (e.g., associative, commutative, distributive properties). Here are some other useful results:

This is how we indicate "union"

If $A \subset B$ then $A \cup B = B$

If $A \subset B$ and $B \subset A$ then $A = B$

If $A \subset B$ then $AB = A$

$A \cup A^c = S$

Probability

These are sets, so we use many of the same definitions (e.g., intersection, union, complement), and all results from set theory apply (e.g., associative, commutative, distributive properties). Here are some other useful results:

If $A \subset B$ then $A \cup B = B$

If $A \subset B$ and $B \subset A$ then $A = B$

If $A \subset B$ then $AB = A$

$A \cup A^c = S$

This is how we indicate "complement"

Probability

These are sets, so we use many of the same definitions (e.g., intersection, union, complement), and all results from set theory apply (e.g., associative, commutative, distributive properties). Here are some other useful results:

If $A \subset B$ then $A \cup B = B$

If $A \subset B$ and $B \subset A$ then $A = B$

If $A \subset B$ then $AB = A$

$A \cup A^c = S$

This is how we indicate "intersection" in probability

Probability

Two definitions where probability theory sometimes uses different terminology than set theory:

A and B are <u>mutually exclusive (disjoint)</u> if they have no outcomes in common.

A and B are <u>exhaustive (complementary)</u> if their union is S.

Probability---definition

We will assign every event A a number P(A), which is the probability the event will occur ($P: S \dashrightarrow R$).

We require that

1. $P(A) >= 0$ for all $A \subset S$
2. $P(S) = 1$
3. For any sequence of disjoint sets $A_1, A_2, \ldots,$

$$P(\cup_i A_i) = \sum_i P(A_i)$$

A <u>probability</u> on a sample space S is a collection of numbers P(A) that satisfy axioms 1-3.

Probability

One can prove a lot of useful things about probabilities using set theory. We'll just state some.

$P(A^c) = 1 - P(A)$

$P(\emptyset) = 0$

If $A \subset B$ then $P(A) <= P(B)$

For all A, $0 <= P(A) <= 1$

$P(A \cup B) = P(A) + P(B) - P(AB)$

$P(AB^c) = P(A) - P(AB)$

Probability

An important special case:

Suppose you have a finite sample space. Let the function $n(.)$ give the number of elements in a set. Then define $P(A) = n(A)/n(S)$. This is called a <u>simple sample space</u>, and it is a probability.

(Check: 1. $P(A)$ will always be non-negative because it's a count. 2. $P(S)$ will equal 1, by definition. 3. $P(A \cup B) = n(A \cup B)/n(S) = n(A)/n(S) + n(B)/n(S) = P(A) + P(B)$.)

Probability

Powerful notion: If you can put an experiment into the framework of a simple sample space (i.e., a sample space where all outcomes are equally likely), all you need to do is count to compute probabilities of events.

What's the probability that the sum of the faces of two fair dice come up 4 when rolled?

$n(S) = 6 \times 6 = 36$ (1&1, 1&2, 1&3, 1&4, 1&5, 1&6, 2&1, . . .)

$n(A) = 3$ (1&3, 2&2, 3&1)

so $P(A) = 3/36 = 1/12$

Probability

If the state of Massachusetts issues 6-character license plates, using one of 26 letters and 10 digits randomly for each character, what is the probability that I will receive an all-digit license plate?

$n(S)$ = 36 possibilities for each of 6 characters = 36^6 = 2.176b

$n(A)$ = 10 possibilities for each of 6 characters = 10^6 = 1m

so $P(A)$ = .0005

This is called sampling with replacement.

Probability

What if Massachusetts does not reuse a letter or digit?

Now, in the sample space, there are 36 possibilities for the 1st character, 35 left for the 2nd, and so on. $n(S) = 36 \times 35 \times 34 \times 33 \times 32 \times 31 = 36!/30!$.

Similarly, in the event, there are 10 possibilities for the 1st character, 9 left for the 2nd, and so on. $n(A) = 10 \times 9 \times 8 \times 7 \times 6 \times 5 = 10!/4!$

so $P(A) = .0001$

This is called <u>sampling without replacement.</u>

Probability

What if Massachusetts does not reuse a letter or digit?

Now, in the sample space, there are 36 possibilities for the 1st character, 35 left for the 2nd, and so on. $n(S) = 36 \times 35 \times 34 \times 33 \times 32 \times 31 = 36!/30!$.

Similarly, in the event, there are 10 possibilities for the 1st character, 9 left for the 2nd, and so on. $n(A) = 10 \times 9 \times 8 \times 7 \times 6 \times 5 = 10!/4!$

so $P(A) = .0001$

We write $n!$ for $n \times (n-1) \times \ldots 3 \times 2 \times 1$

This is called <u>sampling without replacement.</u>

Probability

To compute these probabilities, all I did was count. Some fancy counting, to be sure, but just counting. Here are some rules for fancy counting (combinatorics):

1. If an experiment has two parts, first one having m possibilities and, regardless of the outcome in the first part, the second one having n possibilities, then the experiment has mxn possible outcomes.

Probability

2. Any ordered arrangement of objects is called a <u>permutation</u>. The number of different permutations of N objects is N!. The number of different permutations of n objects taken from N objects is N!/(N-n)!.
3. Any unordered arrangement of objects is called a <u>combination</u>. The number of different combinations of n objects taken from N objects is N!/{(N-n)!n!}. We typically denote this $\binom{N}{n}$ ---"N choose n."

Probability---examples

All candidates for the Republican Presidential nomination gather onstage for an event.

How many handshakes are exchanged if everyone shakes everyone else's hand?

Probability---examples

All candidates for the Republican Presidential nomination gather onstage for an event.
How many handshakes are exchanged if everyone shakes everyone else's hand?

?!

Probability---examples

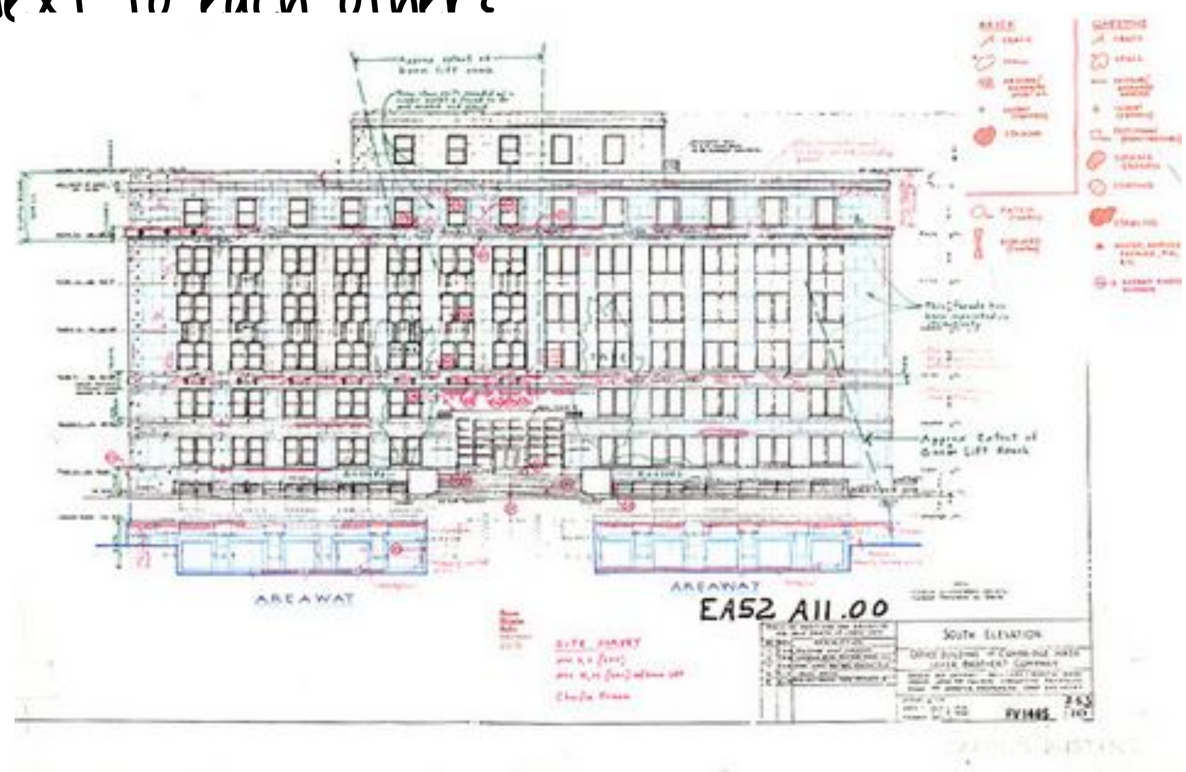
All candidates for the Republican Presidential nomination gather onstage for an event.

How many handshakes are exchanged if everyone shakes everyone else's hand?

$$\binom{9}{2} = 9 \times 8/2$$

Probability---examples

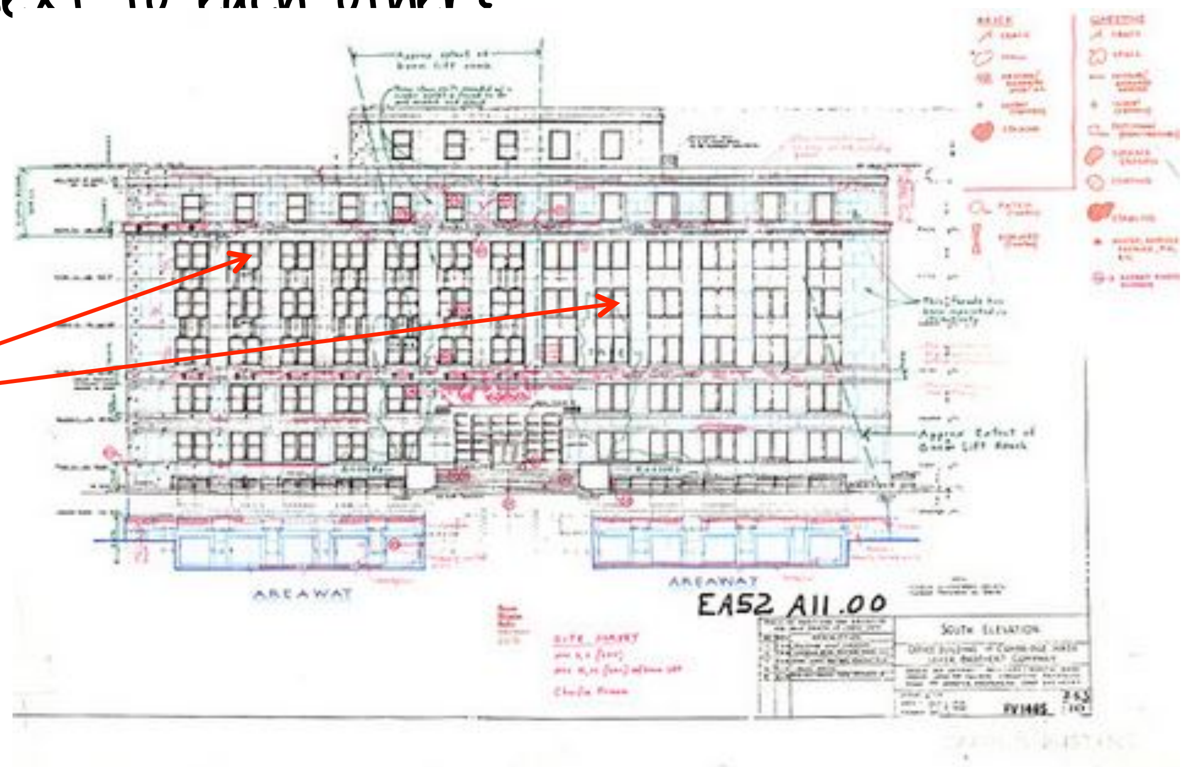
We have 40 faculty offices in the renovated E52. (Assume they're in a continuous line.) If 40 faculty members are placed randomly in the offices, what is the probability that Esther and I are next to each other?



Probability---examples

We have 40 faculty offices in the renovated E52. (Assume they're in a continuous line.) If 40 faculty members are placed randomly in the offices, what is the probability that Esther and I are next to each other?

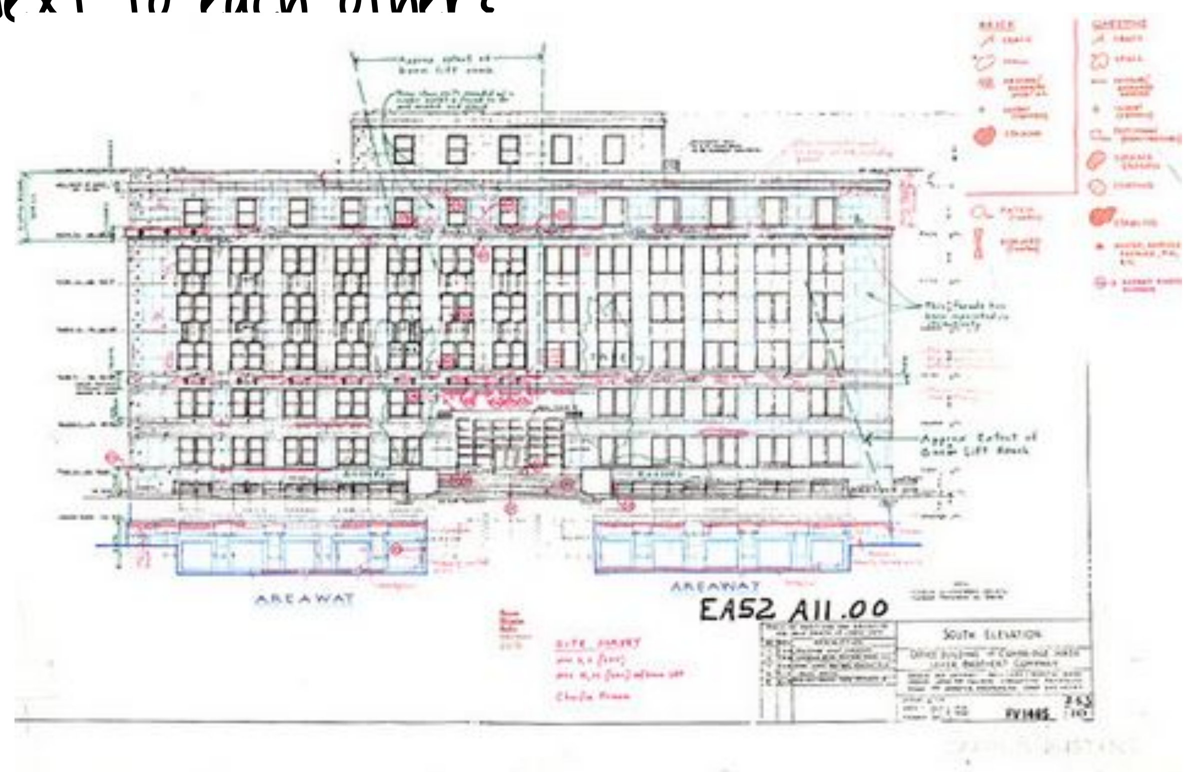
Actually, here we are



Probability---examples

We have 40 faculty offices in the renovated E52. (Assume they're in a continuous line.) If 40 faculty members are placed randomly in the offices, what is the probability that Esther and I are next to each other?

2x39!/40!



Probability---a pre-lunch example

The Area Four menu contains six vegetarian pizza toppings and five non-vegetarian pizza toppings:

EXTRA TOPPINGS

Caramelized Onions, Pickled Banana Peppers, Mushrooms, Green Olives: **$1.50 | $3**

Arugula, Sopressata, Sausage, Bacon, Chicken *: **$2.50 | $4**

2 Farm Eggs *: **$3.5**

Marinated White Anchovies *: **$5/8**

If I write each on a piece of paper and randomly choose two, what is the probability that I end up with a pizza that has one veg and one non-veg topping?

Probability---a pre-lunch example

First characterize the sample space, S:

$S = \{(V1, V2), (V1, V3), (V1, V4), \ldots, (V1, N1), (V1, N2), \ldots\}$ $\quad n(S) = \binom{11}{2} = 55$

(Are all outcomes equally likely? Yes.)

Now characterize A:

$A = \{(V1, N1), (V1, N2), \ldots,$
$(V2, N1), (V2, N2), \ldots,$
$(V3, N1), \ldots\}$ $\quad n(A) = 6 \times 5 = 30$

So the probability is $n(A)/n(S) = 30/55$

Probability---a pre-lunch example

In general, I could have chose n toppings and asked what is the probability that my pizza had n_1 vegetarian toppings and n_2 non-vegetarian toppings. There would, then, be $\binom{6}{n_1}$ possibilities for the veg toppings and $\binom{5}{n_2}$ for the non-veg toppings. In other words,

$$P(n_1 \text{ veg}, n_2 \text{ non-veg}) = \frac{\binom{6}{n_1}\binom{5}{n_2}}{\binom{11}{n}}$$

Probability---a pre-lunch example

In general, I could have chose n toppings and asked what is the probability that my pizza had n_1 vegetarian toppings and n_2 non-vegetarian toppings. There would, then, be $\binom{6}{n_1}$ possibilities for the veg toppings and $\binom{5}{n_2}$ for the non-veg toppings. In other words,

$$P(n_1 \text{ veg}, n_2 \text{ non-veg}) = \frac{\binom{6}{n_1}\binom{5}{n_2}}{\binom{11}{n}}$$

We will refer back to this example as the basis for a special distribution, the hypergeometric.

Probability---independence

It's going to be important for us, going forward, to be able to talk about the relationship between probabilistic, or stochastic, events. The most fundamental of these relationships is independence.

Events A and B are <u>independent</u> if $P(AB) = P(A)P(B)$.

That definition doesn't seem very intuitive, and most of us probably think that we have a good, intuitive sense of what independent events are. Just be careful, though, because that intuition can be misleading.

Probability---independence

Suppose you toss one die. Consider the event, A, that you roll a number less than 5, and the event, B, that you roll an even number. Are these events independent? (How could they be?---they rely on the same roll of a die.)

Yes, they are. Let's check: $P(A) = 2/3$. $P(B) = 1/2$. $P(AB) = 1/3$. (AB is rolling an even number less than 5, i.e., 2 or 4.) and $P(A)P(B) = P(AB)$. (The proper intuition about independent events is that knowing one event occurred doesn't give you any information about whether the other occurred.)

Probability---independence

<u>Thm</u> If A and B are independent, A and B^c are also independent.

<u>Pf</u> $P(AB^c) = P(A)-P(AB) = P(A) - P(A)P(B) = P(A)(1-P(B)) = P(A)P(B^c)$

For more than two events, we define independence the same way---the events are independent if the probability of their intersection is equal to the product of their probabilities.

Probability---example

Probability---example

Steph Curry's 3pt FG percentage is 44%. (Assume independence of shots.)

What is the probability that he misses the next three shots he takes, and then makes the three after that?

Probability---example

What is the probability that he misses the next three shots he takes, and then makes the three after that?

P(miss)P(miss)P(miss)P(make)

P(make)P(make) = $.56^3 \times .44^3$

= .015

(same as *any particular* sequence of 3 misses and 3 shots made---order won't matter.)

Probability---example

What is the probability that he misses three and makes three of the next six shots he takes?

Probability---example

What is the probability that he misses three and makes three of the next six shots he takes?

Just multiply the probability of any one such sequence (.015) by the number of such sequences ($\binom{6}{3}$ = 20), and that equals .30.

Probability---example

What is the probability that he makes at least one shot in the next six he takes?

Probability---example

What is the probability that he makes at least one shot?

Well, certainly could calculate probability that he makes one, the probability he makes two, etc., and add those. There's an easier way:

P(making at least one shot) = 1- P(not making any) = $1 - .56^6$ = .969.

Probability---example

What is the probability that he makes at least one shot?

Well, certainly could calculate probability that he makes one, the probability he makes two, etc., and add those. There's an easier way:

P(making at least one shot) = 1- P(not making any) = $1 - .56^6$ = .969.

We will refer back to this example as the basis for a special distribution, the binomial.

Probability---conditional probability

Recall that knowing that two events are independent means that the occurrence (or nonoccurrence) of one event doesn't tell you anything about the other.

But what if we have two events where the occurrence of one event actually tells us something relevant about the probability of another event? How can we alter the probability of the second event appropriately?

The <u>probability of A conditional on B</u>, $P(A|B)$, is $P(AB)/P(B)$, assuming $P(B) > 0$.

Probability---conditional probability

Recall that knowing that two events are independent means that the occurrence (or nonoccurrence) of one event doesn't tell you anything about the other.

But what if we have two events where the occurrence of one event actually tells us something relevant about the probability of another event? How can we alter the probability of the second event appropriately?

The <u>probability of A conditional on B</u>, $P(A|B)$, is $P(AB)/P(B)$, assuming $P(B) > 0$.

Think about redefining both the event and sample space based on new information.

Probability---conditional probability

What is the relationship between independence and conditional probability?

Suppose A and B are independent and $P(B) > 0$. Then, $P(A|B) = P(AB)/P(B) = P(A)P(B)/P(B) = P(A)$.

This is consistent with our intuition---B occurring tells us nothing about and probability of A, so the conditional probability equals the unconditional probability. (Note that the implication goes both ways: $P(A|B) = P(A)$ iff A & B independent.)

Probability---example

An interesting part of the American political process is the tension between winning over party faithful to get the nomination and being able to appeal to a broader base of voters in the general election.

Let's suppose these candidates have following probabilities of winning the nomination:

Trump	$P(A_1) = .4$
Cruz	$P(A_2) = .3$
Rubio	$P(A_3) = .2$
Carson	$P(A_4) = .1$

Probability---example

Let's suppose that, conditional on winning the nomination, these candidates have following probabilities of winning the general election:

Trump	$P(W\|A_1) = .25$
Cruz	$P(W\|A_2) = .2$
Rubio	$P(W\|A_3) = .6$
Carson	$P(W\|A_4) = .4$

Probability---example

Let's suppose that, conditional on winning the nomination, these candidates have following probabilities of winning the general election:

Trump	$P(W\|A_1) = .25$
Cruz	$P(W\|A_2) = .2$
Rubio	$P(W\|A_3) = .6$
Carson	$P(W\|A_4) = .4$

Tension embodied in fact that candidates with higher probability of winning nomination might not have higher probability of winning general election.

Probability---example

Let's suppose that, conditional on winning the nomination, these candidates have following probabilities of winning the general election:

Trump	$P(W\|A_1) = .25$
Cruz	$P(W\|A_2) = .2$
Rubio	$P(W\|A_3) = .6$
Carson	$P(W\|A_4) = .4$

How can we compute the probability of a Republican win in the general election, $P(W)$?

Probability---example

Let's do a little side calculation:

$P(W) = P(WS)$

$= P(W(A_1 \cup A_2 \cup A_3 \cup A_4))$ because A_1-A_4 are mutually exclusive and exhaustive sets, a <u>partition</u>

$= P(WA_1 \cup WA_2 \cup WA_3 \cup WA_4)$

$= P(WA_1) + P(WA_2) + P(WA_3) + P(WA_4)$

$= P(W|A_1)P(A_1) + P(W|A_2)P(A_2) + P(W|A_3)P(A_3)$
$+ P(W|A_4)P(A_4)$

Probability---example

So, we just plug in to calculate the probability of a Republican win in the general election.

$P(W) = .4x.25 + .3x.2 + .2x.6 + .1x.4 = .32$

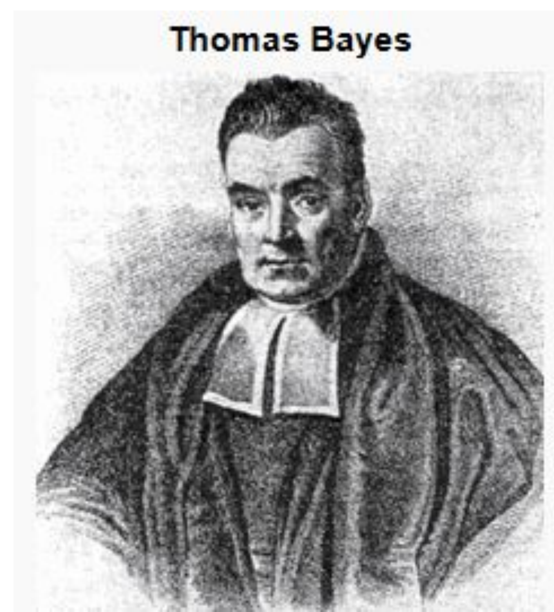
Probability---Bayes' Theorem

We've seen $P(AB) = P(B|A)P(A) = P(A|B)P(B)$ (provided $P(A) > 0$ and $P(B) > 0$), so can write $P(A|B) = P(B|A)P(A)/P(B)$.

We've also seen $P(B) = P(B|A)P(A) + P(B|A^c)P(A^c)$.

So, $P(A|B) = P(B|A)P(A)/\{P(B|A)P(A) + P(B|A^c)P(A^c)\}$.

(A & A^c form a partition of S.
You can do this with any partition of S.)



Probability---example

A pregnant woman lives in an area where the Zika virus is fairly rare---1 in 1000 people have it. Still, she's concerned, so she gets tested. There is a good but not perfect test for the virus---it gives a positive reading with probability .99 if the person has the virus and a positive reading with probability .05 if the person does not. Her reading is positive.

How concerned should she be now?

Probability---example

We can use Bayes' Theorem to calculate the probability she actually has the virus conditional on her positive test.

$P(Z) = .001$ (unconditional probability of having Zika)

$P(Z^c) = .999$

$P(+|Z) = .99$

$P(+|Z^c) = .05$

$P(Z|+) = P(+|Z)P(Z)/\{P(+|Z)P(Z) + P(+|Z^c)P(Z^c)\}$

$= .019$---less than 2% probability!!

Surprising?

Probability---example

How can that be?

The unconditional (prior) probability of her having the virus was quite low, 1/1000. We updated the probability based on the results of an imperfect test, but since it's much more likely that the test was wrong (50 out of 1000 people without the virus test positive), our probability gets updated based on the positive test, but it doesn't get updated that much. It goes from .0001 to .019.