Probability (Continued)

NOTE:

Next class is next Thursday October 12

Discussions next week will not meet—instead Xin will have extended office hours on October 10th from 12:20-2:15 in his office Room 206, MCS (111 Cummington Mall)

HW 4 will be posted on the web page soon. It will be due Thursday 12 October in class. I will email you when it is available.

Independence

We say two events A and B are independent if

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

That is, the events A and B are independent if the probability of an outcome being in both A and in B is the product of the probabilities of A and B.

Examples: Dice

Consider a usual 6-sided die on with the numbers 1 to 6 written on the sides. You roll the die to see which side faces up. The sample space is {1,2,3,4,5,6}.

We use the Equally likely outcomes model says that each of these numbers is equally likely, so each has probability 1/6. That is

$P(\{1\})=P(\{2)\}) = \ldots = P(\{6\})=1/6$

ROLL ONE DIE:

Let event A be roll 2 , 4 or 6 (so A={2,4,6}) and event B be roll 3 or 6 (so B={3,6})

So P(A)= 3/6 = ½ and P(B)= 2/6 = 1/3.

Are events A and B independent?

Well P(A and B) = P({6}) = 1/6   and

P(A) x P(B)= (1/2) x (1/3) = 1/6.

So yes! They are independent.

Suppose A is the same even, roll 2, 4 or 6.

Let event C be event roll 2 or 6 (so C={2,6}).

So P(C)= 2/6 = 1/3.

Are A and C independent?

Well P(A and C) = P({2,6})= 2/6=1/3

But P(A) x P(C)= ½ x 1/3 = 1/6.....so NO, A and C are dependent events.

So our definition of "independent" is easy to check if we know the probabilities of events…but does it agree with our idea of what independent should mean?

If we say A is the event "roll even number"

B is the event "roll a multiple of 3"

And

C is the event "roll an even number that isn't 4"

then A and B sound kinda-sorta independent, but then so do A and C, sorta…

This is why we need the precise definition!

When your intuition (might) fail

There are lots of cases where your intuition about probability might not correspond to the calculations…working with independent events is one of the situations where this happens often.

For example, Hurricane Irma may have been “500 year storm”, meaning that a hurricane of that size hits a particular spot, on average, once every 500 years.

That means the "long term frequency" of storms the size of Irma hitting a spot is 1/500. Or the probability that a storm of that size will hit a particular spot in a given year is 1/500.

It is natural to think that once such a storm does hit, then we are "safe" from such storms for a while…maybe not 499 years, but at least for a while.

However, hurricanes are independent of each other, so being hit by one such hurricane, does not change the probability of being hit by another in the same year.

An example with dice:

Recall, if we roll one die the sample space is {1,2,3,4,5,6}, using the equally likely probability probability model, the probability of each particular outcome is 1/6.

Recall, if we roll two dice (say a blue one and a red one), the sample space is much larger. It is {(1,1),(1,2),(1,3),…(4,5),…,(6,6)}. This list contains 36 entries (multiplication principle) and it is easier to organize it geometrically.

Second Die

First Die

1 , 1	1 , 2	1 , 3	1 , 4	1 , 5	1 , 6
2 , 1	2 , 2	2 , 3	2 , 4	2 , 5	2 , 6
3 , 1	3 , 2	3 , 3	3 , 4	3 , 5	3 , 6
4 , 1	4 , 2	4 , 3	4 , 4	4 , 5	4 , 6
5 , 1	5 , 2	5 , 3	5 , 4	5 , 5	5 , 6
6 , 1	6 , 2	6 , 3	6 , 4	6 , 5	6 , 6

Using the Equally likely outcomes model, every square in this box (every outcome in the sample space) has the same probability, 1/36.

In this case, events that depend only on the first die are independent from events that depend only on the roll of the second die.

Second Die

First Die

1 , 1	1 , 2	1 , 3	1 , 4	1 , 5	1 , 6
2 , 1	2 , 2	2 , 3	2 , 4	2 , 5	2 , 6
3 , 1	3 , 2	3 , 3	3 , 4	3 , 5	3 , 6
4 , 1	4 , 2	4 , 3	4 , 4	4 , 5	4 , 6
5 , 1	5 , 2	5 , 3	5 , 4	5 , 5	5 , 6
6 , 1	6 , 2	6 , 3	6 , 4	6 , 5	6 , 6

Some questions we can now answer by counting:

What is the probability of the two die rolling the same number?

6 outcomes both die the same, so

P(both the same) =6/36=1/6

What is the probability of the two die being different? Using a theorem from last time…

1- P(both the same) = 1 -1/6 =5/6

We can visualize the problem of 3 dice as a cube with 1 to 6 along each of the axes. Figuring the probability is the same a figuring a volume in this cube.

But if we have 4 dice, we can't visualize a 4 dimensional box, so we need to generalize our counting technique (that is motivated by the geometry so far).

But even when we can visualize the cube for the sample space of 3 dice rolls, the counting is still difficult sometimes.

For example: If we roll 3 dice (a red, a blue and a yellow die) what is the probability that 2 or more dice are the same?

Well, we know that there are 6x6x6=216 outcomes on our sample space (Multiplication principle).

We need only count the outcomes with 2 or more die the same. Well, we could make a list…

{(1,1,1), (1,1,2),(1,2,1),(2,1,1), (1,2,2), (1,1,3), …

(oops, need (2,1,2) and (2,2,2)…)

Annoying. Need a better way to count.

Take advantage of what we know!

We know that if A is an event and B is the event made up of all outcomes not in A (called the "complementary event") then

$$P(A) = 1 - P(B)$$

Let A be the event of rolling 3 die and getting 2 or more die the same.

The complementary event is…?

The complementary event is rolling three die and getting 3 different numbers. Let's call this even B. So

$$P(A) = 1 – P(B)$$

In order to figure out the probability of getting 2 or more rolls the same when throwing three dice, (that is, P(A)) we can instead figure out the probability of getting all three rolls different (that is P(B)). This turns out to be easier!

New problem

So what is P(B), the probability of getting all three rolls different?

Well, the first roll could be any of the 6 numbers, so there are 6 ways to do the first roll. Once the first roll is done, to be in event B the second roll must be different and there are 5 ways to be different from the first roll. So

6 x 5 =30 ways for the first 2 rolls.

For each of the 30 possible first two rolls, there are only 4 ways to roll the third die so that it is different from the first two (that we know are different from each other), so there are

(6 x 5) x 4= 30 x 4 = 120

outcomes in event B. Since there are 6x6x6=216 possible outcomes rolling 3 die,

P(B)= 120/216 = 0.555…

(or a little more than ½)

Back to original problem

So P(A) the probability of getting 2 or more die the same number when rolling 3 die is

$$P(A) = 1 - P(B) = 1-0.555\ldots=0.444\ldots$$

Or about 0.44 or 44% or 44/100.

Anybody feeling lucky?

I will bet $100 cash money against…

Big Bet!

I will bet $100 cash money against

That there are at least 2 people in this room right now with the same "birth date" (i.e., same month and day of birthday).

(There are about 100 people in the room…will I win?)

Foolish bet?

Well, there are 366 possible days and only 100 people…seems like the chance 2 or more have the same birthdate would be about….

Foolish bet?

Well, there are 366 possible days and only 100 people…seems like the chance 2 or more have the same birthdate would be about…. 1/3?

(Pretty bad odds for a $ 100 dollar bet.)

But wait. It could be that the first two people in the front row have the same birthdate. Or the first and the third or second and third or….

Let's do it right.

Why am I going to win?

Compute the probability that two or more people in the room have the same birth date.

First a simplification step…throw out leap year (sorry leap year…) So 365 days.

Size of the sample space= number of outcomes

365 choices first person , 365 choices second person…..so 365 x 365 for the first two people

…so sample space has

$365 \times 365 \times 365 \times \ldots = 365^{100}$ possible outcomes.

That's a lot of outcomes.

How many outcomes in which two or more people have the same birth date? (I win)

Really the same problem as the three dice..just bigger. Counting the winning outcomes is too hard…but

Number of outcomes for which everyone has a different birth date (I lose) is easy to count—

365 ways for 1st persons birth date

364 for 2nd person, so 365 x 364 for first 2

363 for 3rd person, so 365x364x363 for first 3

…

365-99 for 100th person…so

365x364x363x….x(365-99) outcomes where

I lose.

So probability that I lose is

$365 \times 364 \times 363 \times \ldots \times (365-99) / 365^{100}$

Breaking this up into a product of fractions

$=(365/365) \times (364/365) \times (363/365) \times$

$\ldots \times ((365-99)/365)$

So probability that I lose is

$365 \times 364 \times 363 \times \ldots \times (365-99) / 365^{100}$

Breaking this up into a product of fractions

$=(365/365) \times (364/365) \times (363/365) \times$

$\ldots \times ((365-99)/365)$

$=0.0000003$

(We are multiplying together a lot of numbers that are less than one)

So the probability I will win is

$1-0.0000003 = 0.9999997$ Let’s see…

Outcome? I Win!!

I took a chance, but not much of one!!

If you try this at home, with 23 people the chance of 2 or more with the same birth date is about ½. 30 people it is about 0.7, 40 people is about 0.9 and 50 people is about 0.97.

Independence

The key to computing the probability above was the use of the Equally likely outcome model—we ASSUMED that every date of the year was as likely to be a given person's birthday as every other date….

Whenever you make an assumption BEWARE!

Ask yourself is it justified…

Independence

For example, if you try this with a members of a hockey team, I have seen reports that the probability of having two people with the same birth date is larger.

Turns out birth dates among hockey players are NOT independent. They tend to be born earlier in the calendar year…Why?

Independence

For example, if you try this with a members of a hockey team, I have seen reports that the probability of having two people with the same birth date is larger.

Turns out birth dates among hockey players are NOT independent. They tend to be born earlier in the calendar year…Why?

Because of the cut off dates for youth hockey!!

Independence

The financial crisis of 2008 was partly (even largely, depending who you listen to) brought about because of defaults in home mortgages. While default is always a possibility for a mortgage loan, it was thought that if many loans were “bundled” together, it would be unlikely that many loans in the same bundle (for homes in Florida, Nevada, California, …) would go into default at the same time…

Independence

This is true if the event of default in one part of the country is independent of default in another part of the country.

If probability of default in Florida is 0.1 and in Nevada is also 0.1 then the probability of default in both at the same time (if they are independent) is $0.1 \times 0.1 = 0.01$ very small!

Independence

BUT

The there are financial shocks that can effect the entire country and cause defaults in different areas to be dependent! In fact, defaults in one stat might even cause defaults in another!

This meant that investments in bundles of mortgages were not as safe as assumed…

The world financial markets almost collapsed.

One more probability problem

At the end of an old TV game show called "Let's make a deal" the host, named Monty Hall, no relation, would offer a contestant a choice of 3 doors. Behind one of the doors was a NEW CAR!!. Behind the other two were goats, literally.

Once the contestant had chosen one of the doors, Monty would have one of the other doors opened. Monty knew which door had the car and which the goats, so he made sure that the door he had opened was a goat.

Then the he would offer the contestant a choice:

Stay with the first choice,

Or

Switch to the remaining closed door?

Monty Hall Problem

The problem is "which is more likely to win—stay or switch?"

The problem is named for Monty Hall (who just passed away). The key to understanding the problem is to remember that Monty knows where the car is. He CHOOSES a goat door to open, so he is giving you information. With more information, you are less uncertain, so the probabilities change!

Glen Hall Problem

Suppose there are 1000 doors with one new car and 999 goats. I will let you pick any door you want. Once you have chosen a door, there are 999 you haven't chosen.

I know where the goats are, so I open 998 of the doors you didn't choose that have goats.

Then I ask if you want to switch from your door to the one other door still closed. Do you switch?

OF COURSE!

In the first choice there was a 1/1000 probability of choosing the car and 999/1000 of getting a goat (so pretty good bet you chose a goat!).

From the remaining 999 doors, I open 998 that have goats, so if you didn't choose the car on your first choice (probability 999/1000), then by switching you are sure to get the car!

So switching gives a probability of winning of 999/1000.

If you don't switch then you don't use the information given by opening doors, so the probability that you win stays the same, 1/1000.

The same is true for 3 doors.

Probability of car in first pick 1/3.

Probability of goat in first pick 2/3.

So probability of car if you switch 2/3.

Probability of car if you don't switch, still 1/3.

Even more detail

Not convinced? You can be even more careful…

Think of the sample space as where the car and goats are placed

Door 1	Door 2	Door 3
Car	Goat	Goat
Goat	Car	Goat
Goat	Goat	Car

We put the equally likely outcome model on these three choices, so each has probability 1/3.

You pick door 1.

For the first outcome, Monty opens door 2 or 3—

You win if you stay, but lose if you switch.

For the second outcome Monty open door 3—

You win if you switch, but lose if you stay

For the third outcome Monty open door 2

You win if you switch, but lose if you stay

Door 1	Door 2	Door 3	Opens	Switch	Stay
Car	Goat	Goat	2 or 3	L	W
Goat	Car	Goat	3	W	L
Goat	Goat	Car	2	W	L

So switching gives you a 2/3 probability of winning while staying with the same door you chose gives you a 1/3 probability of winning.

Door 1	Door 2	Door 3	Opens	Switch	Stay
Car	Goat	Goat	2	L	W
Goat	Car	Goat	3	W	L
Goat	Goat	Car	2	W	L

Remember

Probability is a way to quantify uncertainty.

If there were no uncertainty then every event would have probability 1 or 0. You would know the event will happen or it won't.

The more you know about a situation, the closer the probabilities of events should be to 1 or 0.

Application of Probability

One of the most powerful applications of probability is the field of statistics.

What is “statistics” in a short sentence?

Application of Probability

One of the most powerful applications of probability is the field of statistics.

What is "statistics" in a short sentence?

Statistics is the mathematics of getting information from data.

Polls

We recently experienced a situation in which “statistics failed”.

Almost all polls I saw predicted Clinton would easily win the presidency, but then she didn’t.

Nothing was wrong with the mathematics… really! My faith in mathematics is not shaken!

Template for Doing Mathematics

Problem

|

Model--------------Repeat---------------------|

| Modify

Examples/Conjectures Model

| |

Proof---Did we answer the question?---No

Yes—Fame + $$$

Problem

Predict the outcome of the election using data from a relatively small subset of the total population.

Nothing wrong with that!

We can ask any question we want!

Examples/conjectures/proofs

Examples and conjectures lead us to make precise statements that we can (sometimes) prove.

A proof is an explanation of why a statement MUST be true …….

Examples/conjectures/proofs

Examples and conjectures lead us to make precise statements that we can (sometimes) prove.

A proof is an explanation of why a statement MUST be true ***within the model***.

Two possible ways a bad prediction could be made

We will see that all predictions using statistics are "probabilistic".

"With 95% probability, we predict"…

(Another good saying—on T shirts at conferences of statisticians:

"Statistics means never having to say you are sure.")

OR

Assumptions and simplifications were made in order to make the situation precise enough to use the proven theorems…

That is: The model was too simple…

People lie.

People don’t answer the phone when pollsters call.

People change their minds.

Not everybody who says they will vote actually votes (people lie).

Give up?

NO!

Make a better model and try again!

On to Statistics

No class on Tuesday because Tuesday is Monday.

Next class Thursday which is Thursday

(I will email you when HW 4 is posted)