MA/CS 109 Lecture 9

Probability

(Remember: Homework Due here Thursday)

Probability?

Everybody thinks they understand probability…

We think we have an intuitive feel for how likely it is that some particular thing will happen.

But

But, actions speak louder than words—

People build houses in flood plains—

People buy lottery tickets—

People are “sure” the next coin flip will be heads because the last three were tails and so heads is “due”…

What do we mean by "probability"

Likelyhood of some particular thing happening.

Fraction--# ways to win/ all "outcomes"

Ratio-- """" ""

Coin always heads or tails 50/50 chance of one or the other "long term frequency"

Probability

The canonical example…Suppose I flip a coin…

If the coin is a “fair” coin, then we say that the probability of heads is 1/2 .

And the probability of tails is 1/2 .

Everybody knows this…

But what does this mean?

This is not as easy a question as might first appear.

The flight of the coin is governed by the laws of physics, just like everything else.

If we know the initial velocity when the coin leaves my fingers and where it will be caught, then we should be able to predict exactly if it come up heads or tails…flipping a coin is a "deterministic" process.

In fact, any good magician can flip a coin so that it always comes up heads or always comes up tails.

But for most of us, we have no way to predict if the coin will be heads or tails. This is where probability comes in—

Probability gives us a way to

In fact, any good magician can flip a coin so that it always comes up heads or always comes up tails.

But for most of us, we have no way to predict if the coin will be heads or tails. This is where probability comes in—

Probability gives us a way to

"Quantify our uncertainty"

For the coin flip—we know for sure that the coin will come up either heads OR tails (it won't be allowed to land on its edge) but we have no information telling us which is more likely, so both heads and tails have the same probability of 1/2.

For the lottery, we do not know if we will win the big prize or not, but we know that winning the big prize is very small…1/400,000,000, say, while the chance of not winning the big prize is 399,999,999/400,000,000.

Interpretation of Probability numbers

One way to interpret these numbers is as “long-term frequency” If we flip a coin many hundreds of time, then we expect half the time it will be heads and half the time it will be tails. If we play the lottery many, many trillions of times, we expect to win 1 out of every 400,000,000 times and lose the rest.

Probability

So probability is a way to assign a number to something that might or might not happen to help us judge how likely that thing might be—what its long range frequency is.

We think of this number as the number of times the particular event will occur if we repeat the same experiment in exactly the same way many, many times.

Probability

We use probability when our knowledge is insufficient to know for sure what will happen. If our knowledge increases, then we will change the probability.

But always we are "quantifying our uncertainty".

Probability and Statistics

Probability and Statistics are part of the mathematical sciences. At BU we are the "Department of Mathematics and Statistics" and a number of my colleagues list "Probability" as their research area (like "geometry" or "algebra"…)

But it is also a huge field (the BU Medical School has a Department of Biostatistics).

We use the same "Template…"

Template for Doing Mathematics

```
Problem
   |
Model--------------Repeat---------------------|
   |                                        Modify
Examples/Conjectures                        Model
   |                                          |
Proof---Did we answer the question?---No
                  Yes—Fame + $$$
```

Problem

Our problem is "How do we talk about and compare the chance of an event happening in a quantitative way."

That is

"How do we quantify our uncertainty."

And the first job is to build a model framework for doing this.

Vocabulary

As always in mathematics, we require that everyone understand the words we use (ourselves included!). This means making precise definitions of terms.

This might make different areas of the mathematical sciences seem a bit mysterious before you have learned the words.

Outcome = possible result of an experiment or action. For example, when flipping a coin, the outcomes are Head (H) and Tails (T).

Sample space = the set of all possible outcomes of an experiment or action. For coin flip the sample space is {H,T} since these are the only things that can happen. If we flip a red and a blue coin the sample space is (red listed first) {(H,H), (H,T), (T,H), (T,T)}

Event = Subset of the possible outcomes. So for flipping a coin once, {H} is an event. For flipping a coin twice getting H first is an event, the event {(H,H), (H,T)}

Don't be intimidated by the vocabulary (…and don't use it to intimidate others).

Probability is a way to assign numbers to events. That is, it is a function that takes events to numbers.   A probability function must satisfy the following rules or axioms:

1. Every event is assigned a number between 0 and 1 (with both 0 and 1 possible).
2. The probability of the event that is the entire sample space is 1.

These axioms are completely reasonable.

For axiom 1: If a probability is supposed to represent the fraction of the times a particular event occurs if an experiment is repeated many times, then the largest it can be is 1 (always happens) and the smallest it can be is 0 (never happens). Everything else is in between 0 and 1.

For axiom 2: The entire sample space is, by definition, the set of every possible outcome (everything that can happen). So what happens when we do the experiment is one of the outcomes, the probability of that thing being in the sample space is 1.

We only need one more axiom, but it is a bit more complicated

3. If two events do not share any outcomes, then the probability of being in either event is the sum of the probabilities of the individual events.

If we let A and B be events (subsets of the sample space) and we know that A and B are "disjoint" (have no common members), then the probability of an experiment giving an outcome in A or B is the probability of getting an outcome in A plus the probability of getting an outcome in B.

Axiom 3

The event made up of all the outcomes in A along with all the outcomes in B is called "A or B" (the outcomes that are in A or in B).

So Axiom 3 says if A and B are disjoint events then the probability of the event (A or B) is the sum of the probability of A and the probability of B.

Notation:

We can't get farther with long sentences like this—we need a notation that allows us to write more efficiently.

Let S denote our sample space (everything that can happen). Let A,B,C,… denote events (subsets of outcomes in S).

The probability of A is denoted $P(A)$.

Notation-continued

So P(A) is the **probability function** assigning numbers between zero and 1 to events.

Depending on the situation, we might list P(A) as a fraction (2/3) or a decimal 0.666…, whichever is more convenient.

Axiom 1 says: For any event A, P(A) is between 0 and 1 (with 0 and 1 possible).

Axiom 2 says: For the event S (the entire sample space=everything that can happen), P(S)=1.

Axiom 3 says: If A and B are disjoint events (there are no outcomes that are in both A and B) then P(A or B) = P(A) + P(B) .

Interpretation of Axiom 3

Axiom 3 is a little more technical, but it still agrees with our intuition of what probability should mean. It basically says that making an event bigger (for example, adding more ways to "win") increases the probability of that event (makes it more likely you'll win). The increase will be the probability of the added outcomes.

Model for Probability

These axioms are part of the “model” step for addressing the problem of quantifying uncertainty.

The axioms will be satisfied for any function we call a “probability function”. This determines some properties (but not all values) of probability functions.

For example:

Conjecture: If A is an event in S and B is the set of all outcomes that are not in A, then

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

Conjecture: If A is an event in S and B is the set of all outcomes that are not in A, then

$$P(B)= 1 – P(A).$$

Proof: We know P(S)=P(sample space)=1. Also we know every outcome is in A or B. So the event (A or B)=S, the event (A or B) is the entire sample space. But A and B are disjoint, so

P(A or B) =P(A) + P(B).

And we know P(A or B) = P(S) =1.

So 1=P(A) + P(B).

So 1-P(A)=P(B).

Theorem: If A is an event in S and B is the set of all outcomes that are not in A, then

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

We know now that this is true for ANY probability function…it isn't an axiom but it follows from just the axioms.

Here's another example:

Conjecture: Let the “empty set” be the event containing no outcomes. Then

P(empty set) = 0

Proof: Know P(S)=1. S event contains all outcomes. So empty set contains all outcomes not in S. So S is S or empty set, and empty set and S are disjoint. Theorem says

P(S)+P(empty set) =P(S)

Or 1+P(empty set) =1. or P(empty set)=0

Theorem: Let the “empty set” be the event containing no outcomes. Then

P(empty set) = 0

Proof: We know P(S)=P(sample space)=1

The empty set is the set containing no outcomes, so the empty set and S are disjoint.

So P(S) + P(empty set)=P(S or empty set)=P(S)

So 1 + P(empty set) = 1

Or P(empty set)=1-1=0.

Choosing the Probability Function

To complete our model of a particular situation, we still have to choose the probability function for all events not equal to S or the empty set.

Of course, it must satisfy the axioms (in order to even be called a probability function and in order to work like we think probabilities should work).

Exactly what probability function we choose will depend completely on the details of the situation we are modeling.

The “Coin Toss”

Suppose we toss a coin. The outcomes are “heads” (we’ll write “H”) and “tails” (we’ll write “T”).

The sample space is S={H,T} and the possible events are

Empty set, {H}, {T}, {H,T}

To define a probability function, we already know we must have P(Empty set)=0 and P({H,T})=P(S)=1.

So there are only two numbers left to choose….

The "Coin Toss"

To satisfy axiom 3, we must have

$$P(\{H\})+ P(\{T\}) = 1$$

(because the event ({H} or {T})={H,T}=S).

So

$$P(\{H\})= 1- P(\{T\})$$

This still leaves lots of possibilities:

Probability functions for coin toss

"Fair Coin": For the fair coin, we have no reason to believe or suspect that H will come up more or less often than T, so we must assign them equal probabilities or

$$P(\{H\})=P(\{T\}) = 1/2 .$$

"Double headed coin": For a coin with "Heads" or H on both sides, T can never appear so we

$$P(\{T\}) = 0 \text{ and } P(\{H\}) = 1.$$

These are both legitimate probability functions, both satisfy the 3 axioms (as you can, and should check). But they correspond to very different situations.

So choosing the probability functions gives us tremendous flexibility to study different situations.

Most Commonly Used Probability Function

While the choice of probability function should be adapted to the situation and knowledge you have, the most commonly used probability function is called the

“Equally likely outcomes model”

It is used in many gambling games because it embodies our idea of “fair” for such games. It also reflects the situation when we have no knowledge of which outcome is more likely than any other…

Equally likely outcomes model

Suppose our sample space contains only finitely many outcomes (so there are only finitely many different things that can possibly happen).

Then every event must also contain only finitely many possible outcomes (since an event is a subset of the sample space).

The Equally Likely Outcomes Model assigns probability values to events as follows: If A is an event then

$$P(A) = \frac{(\text{Number of outcomes in A})}{(\text{Number of outcomes in S})}$$

That is, we determine the value of the probability of A by counting the number of outcomes in the event A and dividing by the number of outcomes in all of S (which is the number of possible outcomes).

We need to check that this assignment of probability values satisfies our 3 axioms.

Axiom 1 is satisfied because the number of outcomes in any event is less than or equal to the total number of possible outcomes, so

$$\frac{\text{number of outcomes in event}}{\text{total number of possible outcomes}}$$

Must be between 0 and 1.

Axiom 2 is OK too—the probability of the whole sample space S is

$$P(S) = \frac{\text{number of outcomes in the sample space}}{\text{number of outcomes in the sample space}}$$

So P(S)=1.

Verifying the third axiom is a little trickier—

If A and B are two disjoint events, then the number of outcomes in (A or B) is the sum of the sum of the outcomes in A plus the outcomes in B, So

$$P(A \text{ or } B) = \frac{\#\text{ outcomes A or B}}{\#\text{ outcomes in S}}$$

$$= \frac{\#\text{ outcomes in A} + \#\text{ outcomes in B}}{\#\text{ outcomes in S}}$$

$$= \frac{\#\text{ outcomes in A}}{\#\text{ outcomes in S}} + \frac{\#\text{ outcomes in B}}{\#\text{ outcomes in S}}$$

$$= P(A) + P(B)$$

So the Equally likely outcomes model is a valid way to assign probabilities to events.

If an event contains just one outcome, then the probability of that event is

P(event with one outcome)

=1/(number of outcomes in S)

If an event has two outcomes, then we can express it as the event with one of the outcomes OR the event with the other outcome.

Axiom 3 says that the probability of the event with 2 outcomes must be

P(two outcomes) =P(one outcome)
+P(other outcome)
=1/(# of outcomes in S)
+1/(# of outcomes in S)

=2/(# of outcomes in S).

And so on…

One advantage of the Equally likely outcomes model is that the determination of P(A) for any event A comes down to counting the elements of A…

Counting is something we have been doing a long time, and we are good at it!

Another advantage is that Equally likely outcomes seems “fair”. Intuitively, it says no outcome will turn up more or less often, over the long term, than any other outcome.

But equally likely outcomes is not always the “right” model…in Boston, most days are not rainy. In fact, “weather-and-climate.com” says that on average there are 128 rainy days in Boston per year (days on which it rains at least a little).

So if our sample space {rains in Boston, doesn’t rain in Boston} then we should set

P(rains in Boston)=128/365=0.35

P(doesn’t rain in Boston)=1-0.35 = 0.65

Not equally likely outcomes…

Another example:

We often hear the phrase "everyone should have an equal chance to get ahead". If the sample space is the set of all possible financial lives of each person, then the statement above is really the equally likely outcome model.

Most politicians would say that the probability of the event "person X will become rich" should be the same for every person, at least at the moment they are born.

But experience says that is not true, the probability "person X will become rich" is much higher if X has rich parents.

But imagine if a politician said

"The probability that a baby born today will become rich in their lifetime is not the same for every baby. My platform is that the probability of becoming rich should at least be non-zero for every baby."

They wouldn't stand much chance of being elected…(Jimmy Carter once famously said

"Life isn't fair.")

The Equally likely outcome model is a good model for card and dice games because we believe each card or each possible roll of the dice is equally likely (at least in “fair” games), so they make good examples for probability.

But card and dice games usually involve many repeats of the same action (dealing many cards or rolling the dice many times). We need one more idea before doing examples.

This idea is also tied up with our notions of “fair”.

Independent Events

What should it mean to say that two events are "independent":

Mutually exclusive

Don't effect each other…

Probability of one event doesn't effect probability of the other.

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.

Our definition is that A and B are independent events if

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

Does this agree with our intuition??

It does say that the probability of getting an outcome in both events depends only on the probability of each of the events by itself.

But why this formula?

Visualization of Independence

Suppose the sample space is all points in a square

The probability of an event is the fraction of the area of the box that the event covers.

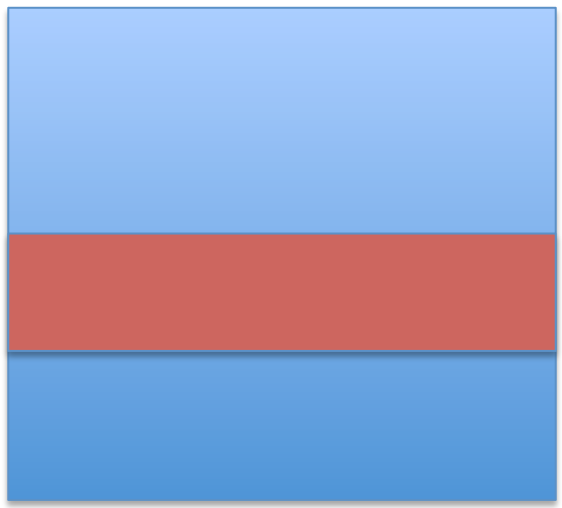
Visualization of Independence

If I choose a point "at random" inside the square

then the location of the point horizontally should be "independent" of its vertical location.

Visualization of Independence

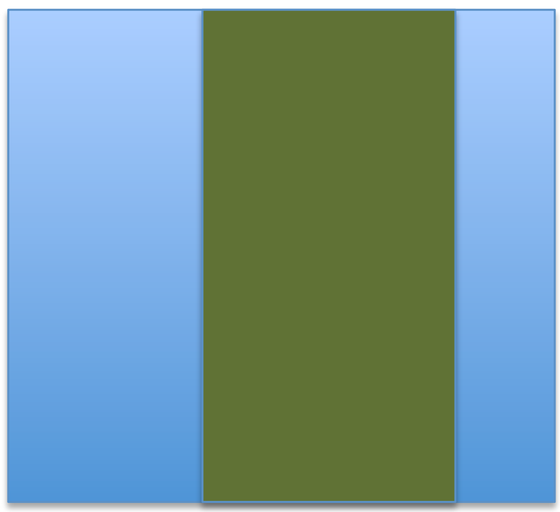
Let A be the event given by the horizontal strip



P(A) is the fraction of the area the strip covers so it is the fraction of the height that it covers.

Visualization of Independence

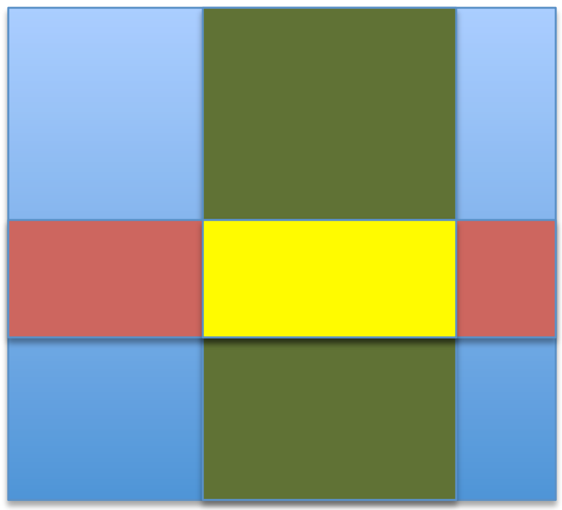
Let B be the event given by the vertical strip



P(B) is the fraction of the area the strip covers so it is the fraction of the width that it covers.

Visualization of Independence

Events A and B are independent because



Because the event (A and B), the yellow box, covers an area given by the width times the height, so P(A and B) = P(A) x P(B).

We say that two events A and B are not independent if the probability of both is not the product of the probabilities, that is

P(A and B) is not P(A) x P(B)

Superstition is really just refusal to believe in the independence of events…If I wear my lucky shirt I will get an A on my math test…We didn't have much snow last year, we are due for a hard winter…

Examples: Coins

Suppose we have a "fair" coin. So the sample space is {H,T} (heads and tails) and P({H})=1/2=P({T}).

If we flip the coin twice, then our sample space becomes larger {(H,H), (H,T), (T,H), (T,T)} (with the first letter the first coin flip).

Saying “the two flips are independent” is saying that

If A is an event that depends only on the first flip and B is an event that depends only on the second flip, then

P(A and B) = P(A) x P(B)

For example if A=first flip heads ={(H,H), (H,T)}

and B = second flip tails= {(H,T),(T,T)} then

Saying the flips are independent is the same as saying that

P(A and B) = P({(H,T)} = P(A) x P(B).

This is true of the Equally likely outcomes model

$P(A \text{ and } B) = P(\{(H,T)\}) = ¼ = ½ \times ½ = P(A) \times P(B)$

But it can be true for other models as well.

Suppose the first coin is fixed so that it always comes up tails and the second coin is fair.

Then $P(\{(H,H)\})=P(\{(H,T)\}) = 0$

While $P(\{(T,H)\})=1/2=P(\{(T,T)\})$

If A=first flip heads ={(H,H), (H,T)}
and B = second flip tails= {(H,T),(T,T)} then
P(A and B) = P({(H,T)}) = 0
But
P(A)= 0 because the first flip is never heads
P(B)= ½.
So P(A and B) = 0 = 0 x ½ = P(A) x P(B)
So A and B are independent!

Saying that the flips are independent says only that they “don’t influence” each other, not that they are the same or “fair”.

On the other hand, suppose a magician is flipping the second coin and always arranges for it to be the same as the first flip (which is fair)…

Then $P(\{H,H\})=1/2$, $P(\{H,T\})=0$, $P(\{T,H\})=0$ and $P(\{T,T\})=1/2$.

If A is the event "first flip heads"

Then $P(A)= P(\{\{H,H\}, \{H,T\}\})=$

$$P(\{H,H\})+P(\{H,T\})=1/2+0=1/2$$

While if B is "second flip tails"

Then $P(B)=P(\{\{H,T\},\{T,T\}\})=$

$$P(\{H,T\})+P(\{T,T\}) = 0 +1/2=1/2$$

But…

P(A and B)

=P(first flip heads and second flip tails)

=P({H,T})= 0

And this is NOT equal to P(A) x P(B)= ½ x ½ = ¼

So for this probability function A and B are dependent (as the situation implies—the flip of the magician depends on the first flip).