Lecture 4 : Conditional Probability and Bayes' Theorem

The conditional sample space Motivating examples

1. Roll a fair die once

$$S = \boxed{\begin{matrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{matrix}}$$

$$\begin{aligned} \text{Let}\quad A &= \text{6 appears} \\ B &= \text{an even number appears} \end{aligned}$$

So

$$P(A) = \frac{1}{6}$$

$$P(B) = \frac{1}{2}$$

Now what about

$$P\left(\begin{array}{c}\text{6 appears } \textit{given} \text{ an even} \\ \text{number appears}\end{array}\right)$$

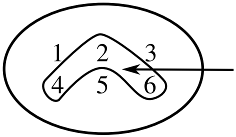
Philosophical Remark

(Ignore this remark unless you intend to be a scientist)
At present the above probability does not have a formal mathematical definition but *we can still compute it*. Soon we will give the formal definition and our computation "will be justified". This is the mysterious way mathematics works. Somehow there is a deeper reality underlying the formal theory.

Back to Stat 400

The above probability will be written given $P(A \text{ given } B)$ to he read $P(A \text{ given } B)$.

Now we know an even number occurred so the sample space changes



So there are only 3 possible outcomes given an even number occurred so

$$P(6 \text{ given an even number occurred}) = \frac{1}{3}$$

The new sample space is called the *conditional sample space*.

2. Very Important example

Suppose you deal two cards (in the usual way without replacement). What is $P(\heartsuit\heartsuit)$ i.e., P(two hearts in a row).

Well, $P(\text{first heart}) = \dfrac{13}{52}$.

Now *what about the second heart*?

Many of you will come up with $^{12}/_{51}$ and

$$P(\heartsuit\heartsuit) = \left(^{13}/_{52}\right)\left(^{12}/_{51}\right)$$

There are *TWO* theoretical points hidden in the formula.
Let's first look at

$$P(\underbrace{\heartsuit \text{ on } 2^{\text{nd}}}_{\text{this isn't really correct}}) = {}^{12}/_{51}$$

What we really computed was the *conditional probability*

$$P(\heartsuit \text{ on } 2^{\text{nd}} \text{ deal} \mid \heartsuit \text{ on first deal}) = {}^{12}/_{51}$$

Why Given we got a heart on the first deal the conditional sample space is the "new deck" with 51 cards and 12 hearts so we get

$$P(\heartsuit \text{ on } 2^{\text{nd}} \mid \heartsuit \text{ on } 1^{\text{st}}) = {}^{12}/_{51}$$

The second theoretical point we used was the formula which we will justify formally later

$$P(\heartsuit\heartsuit) = P(\heartsuit \text{ on } 1^{\text{st}})P(\heartsuit \text{ on } 2^{\text{nd}} \mid \heartsuit \text{ on } 1^{\text{st}})$$
$$= \binom{13}{52}\binom{12}{51}$$

Basic Questions

These examples will occur repeatedly in today's lecture.

1. What is

$$P(\underbrace{\heartsuit \text{ on } 1^{\text{st}} \mid \heartsuit \text{ on } 2^{\text{nd}}}_{\text{reverse of pg. 5}})$$

   and (easier).

2. What is $P(\heartsuit \text{ on } 2^{\text{nd}}$ with no information on what happened on the $1^{\text{st}})$.

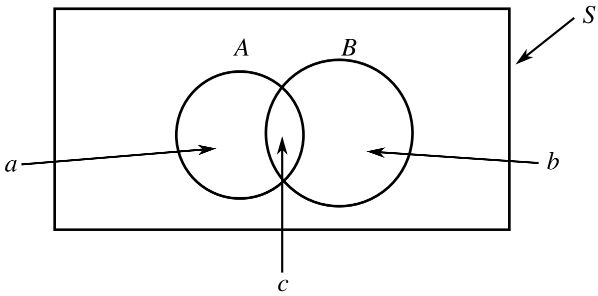
A Second Gold Star Problem

If you do 1- and 2. in the way described on this page you will get a gold star and learn a lot.

Don't use pg. 16 in Bayes Theorem

Let S = set of unordered pairs of distinct cards
Compute $\sharp(S)$.
Let A = subset of pairs of hearts = $\{\heartsuit\heartsuit\} \subset S$.
Compute $\sharp(A)$ and $P(A) = \dfrac{\sharp(A)}{\sharp(S)}$.
Let B = subset of pairs so that the second card is a heart.
Compute $\sharp(B)$ and $P(B)$ solving 2.
Now compute $P(\heartsuit \text{ on } 1^{\text{st}} \mid \heartsuit \text{ on } 2^{\text{nd}})$ by taking the ratio (pg. 11) or else computing the conditional sample space.

The Formal Mathematical Theory of Conditional Probability



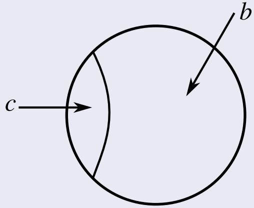
$$\sharp(S) = n,\ \sharp(A) = a,\ \sharp(B) = b,\ \sharp(A \cap B) = C$$

Problem

Let S be a finite set with the equally - likely probability measure and A and B be events with coordinalities shown in the picture.

Problem

Compute $P(A|B)$.
We are given B occurs so the conditional sample space is B



Only part of A is allowed since we know B occurred namely $A \cap B$

$$P(A|B) = \frac{\sharp(A \cap B)}{\sharp(B)}$$
$$= \frac{c}{b}$$

We can rewrite this as

$$P(A|B) = \frac{a}{b} = \frac{a/n}{b/n} = \frac{P(A \cap B)}{P(B)}$$

so

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \qquad (*)$$

This intuitive formula for the equally likely probability measure leads to the following.

Formal Mathematical Definition

Let A and B be any two events in a sample space S with $P(B) \neq 0$. The conditional probability of A given B is written $P(A|B)$ and is *defined* by

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \qquad (*)$$

so if $P(A) \neq 0$ then

$$P(B|A) = \frac{P(B \cap A)}{P(A)} = \frac{P(A \cap B)}{P(A)} \qquad (**)$$

Since $A \cap B – B \cap A$.

We won't prove the next theorem but you could do it and it is useful.

Theorem

Fix B with $P(B) \neq 0$. $P(\cdot|B)$ satisfies the axioms (and theorems) of a probability measure - see Lecture 1.

For example

1. $P(A_1 \cup A_2|B) = P(A_1|B) + P(A_2|B) - P(A_1 \cap A_2|B)$
2. $P(A'|B) = 1 - P(A|B)$

Z $\quad P(A|\cdot)$ *does not satisfy the axioms and theorems.*

The Multiplicative Rule for $P(A \cap B)$

Rewrite (**) as

$$P(A \cap B) = P(A)P(B|A)(\sharp)$$

($\sharp$) is very important, more important then (**).
It complement the formula

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

Now we know how P interacts with the basic binary operations $\cup$ and $\cap$.

More generally

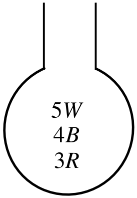
$$P(A \cap B \cap C) = P(A)P(B|A)P(C|A \cap B)$$

Exercise

Write down $P(A \cap B \cap C \cap D)$.

Traditional Example

An urn contains 5 white chips, 4 black chips and 3 red chips.
Four chips are drawn sequentially without replacement. Find $P(WRWB)$.



12 chips

Solution

$$P(WRWB) = \left(\frac{5}{12}\right)\left(\frac{3}{11}\right)\left(\frac{4}{10}\right)\left(\frac{4}{9}\right)$$

What did we do formally

$$P(WRWB) = P(W) \cdot P(R|W)$$

$P(W|W \cap R) \cdot P(B|W \cap R \cap W)$

Now we redo the gold star problem (not the gold star way) namely we compute

$$P(\underbrace{\heartsuit \ \ on \ \ 1^{st}}_{A} \mid \underbrace{\heartsuit \ \ on \ \ 2^{nd}}_{B})$$

By Definition

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(\heartsuit\heartsuit)}{P\left(\begin{array}{c}\heartsuit \text{ on } 2^{\text{nd}} \text{ with the} \\ \text{other information}\end{array}\right)}$$

Now we know from pg. 5.

$$P(\heartsuit\heartsuit) = (13/52)(12/51)$$

Now we need

$$P(\heartsuit \text{ on } 2^{\text{nd}} \text{ with no other information}) = 13/52$$

We will prove this on pg. 18 - it also follows from the gold star approach on pg. 7. So

$$= P(\heartsuit \text{ on } 1^{\text{st}} \mid \heartsuit \text{ on } 2^{\text{nd}}) = \frac{(\cancel{13/52})(12/51)}{(\cancel{13/52})} = \frac{12}{51}$$

$$= P(\underbrace{\heartsuit \text{ on } 2^{\text{nd}} \mid \heartsuit \text{ on } 1^{\text{st}}}_{\text{pg. 5}})!!$$

Bayes' Theorem (pg. 72)

Bayes' Theorem is a truly remarkable theorem. It tells you "how to compute $P(A|B)$ if you know $P(B|A)$ and a few other things".
For example - we will get a new way to compute are favorite probability $P(\heartsuit \text{ as } 1^{\text{st}} \mid \heartsuit \text{ on } 2^{\text{nd}})$ because we know $P(\heartsuit \text{ on } 2^{\text{nd}} \mid \heartsuit \text{ on } 1^{\text{st}})$.
First we will need on preliminary result.

The Law of Total Probability

Let $A_1, A_2, \ldots, A_k$ be mutually exclusive
($A_i \cap A_j = \emptyset$) and exhaustive.
($A_1 \cup A_2 \cup \ldots \cup A_k = S =$ the whole space)
Then for any event B

$$P(B) = P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + \cdots + P(B|A_k)P(A_k) \tag{b}$$

Special case $k = 2$ so we have A and A'

$$P(B) = P(B|A)P(A) + P(B|A')P(A') \tag{bb}$$

Now we can *prove*
$P(\heartsuit$ on 2^{nd} with no other information$) = {}^{13}/_{52}$

$$\text{Put} \quad B = \heartsuit \ \text{ on } \ 2^{\text{nd}}$$
$$A = \text{heart on } \ 1^{\text{st}}$$
$$A' = \text{a nonheart on } 1^{\text{st}}$$

Lets write $\cancel{\heartsuit}$ for nonheart.
So,

$$P(\cancel{\heartsuit} \ \text{ on } \ 1^{\text{st}}) = {}^{39}/_{52}$$
$$P(\heartsuit \ \text{ on } \ 2^{\text{nd}} / \cancel{\heartsuit} \ \text{ on first}) = {}^{13}/_{51}$$

Now

$$
\begin{aligned}
P(B) &= P(B|A)P(A) + P(B|A')P(A') \\
&= P(\heartsuit \text{ on } 2^{\text{nd}} \mid \heartsuit \text{ on } 1^{\text{st}})P(\heartsuit \text{ on } 1^{\text{st}}) \\
&\quad + P(\heartsuit \text{ on } 2^{\text{nd}} \mid \cancel{\heartsuit} \text{ on } 1^{\text{st}})P(\cancel{\heartsuit} \text{ on } 1^{\text{st}}) \\
&= (12/51)(13/52) + (13/51)(39/52)
\end{aligned}
$$

add fractions

$$
= \frac{(12)(13) + (13)(39)}{(51)(52)}
$$

factor out 13 add to get 51

$$
= \frac{(13)(12 + 39)}{(51)(52)} = \frac{(13)(\cancel{51})}{(\cancel{51})(52)}
$$

$= (13)/(52)$ Done!

(b) is "very easy" to prove but we won't do it.
Now we can state Bayes' Theorem.

Bayes' Theorem (pg. 73)

Let $A_1, A_2, \ldots, A_k$ be a collection of n mutually exclusive and exhaustive events with $P(A_i) > 0$
$i = 1, 2, \ldots, k$. Then for any event B with $P(B) > 0$

$$P(A_j|B) = \frac{P(B|A_j)P(A_j)}{\sum\limits_{i=1}^{k} P(B|A_i)P(A_i)}$$

Again we won't give the proof.

Special Case $k = 2$

Suppose we have two events A and B with $P(A) > 0$, $P(A') > 0$ and $P(B > 0)$. Then

$$\sharp) P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A')P(A')}$$

Now we will compute (for the last time)

$$P(\heartsuit \text{ on } 1^{\text{st}} \mid \heartsuit \text{ on } 2^{\text{nd}})$$

Using Bayes' Theorem.
This is the obvious way to

do it since we know the probability “the other way around”

$$P(\heartsuit \text{ on } 2^{\text{nd}} \mid \heartsuit \text{ on } 1^{\text{st}}) = {}^{12}/_{51}$$

So let’s do it.

$$\begin{aligned} \text{We put } A &= \heartsuit \text{ on } 1^{\text{st}} \\ \text{so } A' &= \not\heartsuit \text{ on } 1^{\text{st}} \\ \text{and } B &= \heartsuit \text{ on second} \end{aligned}$$

plugging into ($\sharp$) we get

$$P(\heartsuit \text{ on } 1^{\text{st}} \mid \heartsuit \text{ on } 2^{\text{nd}})$$

$$= \frac{P(\heartsuit \text{ on } 2^{\text{nd}} \mid \heartsuit \text{ on } 1^{\text{st}})P(\heartsuit \text{ on } 1^{\text{st}})}{P(\heartsuit \text{ on } 2^{\text{nd}} \mid \heartsuit \text{ on } 1^{\text{st}})P(\heartsuit \text{ on } 1^{\text{st}}) + P(\heartsuit \text{ on } 2^{\text{nd}} \mid \not\heartsuit \text{ on } 1^{\text{st}})P(\not\heartsuit \text{ on } 1^{\text{st}})}$$

$$= \frac{(^{12}/_{51})(^{13}/_{52})}{(^{12}/_{51})(^{13}/_{52}) + (^{13}/_{51})(^{39}/_{52})}$$

swap numbers use

$$\left(\frac{a}{b}\right)\left(\frac{c}{d}\right) = \left(\frac{c}{b}\right)\left(\frac{a}{d}\right)$$

something from high school

$$= \frac{(^{13}/_{51})(^{12}/_{52})}{(^{13}/_{51})(^{12}/_{52}) + (^{13}/_{51})(^{39}/_{52})}$$

factor this out

$$= \frac{(^{13}/_{51})(^{12}/_{52})}{(^{13}/_{51})\left(^{12}/_{52} + \frac{39}{52}\right)}$$

$$= \frac{\cancel{(^{13}/_{51})}(^{12}/_{52})}{\cancel{(^{13}/_{51})}(^{51}/_{52})} = \frac{(^{12}/\cancel{_{52}})}{(^{51}/\cancel{_{52}})}$$

$= {}^{12}/_{51}$ once again!

The algebra was hard but the approach was the most natural - a special case of

General Principle

If you know $P(B|A)$ and you want to compute $P(A|B)$ use Bayes' Theorem in the ($\sharp$) version, pg. 22.

Compulsory Reading (for your own heath)

In case you or someone you love tests positive for a *rare* (this is the point) disease, read Example 2.30, pg. 73. Misleading (and even bad) statistics is rampant in medicine.