LECTURE 4

Conditional Probability and Bayes' Theorem

1 The conditional sample space

Physical motivation for the formal mathematical theory

1. Roll a fair die once so the sample space S is given by

$$S = \{1, 2, 3, 4, 5, 6\}.$$

Let

$$A = 6 \text{ appears}$$

$$B = \text{an even number appears,}$$

so

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

$$P(A) = \tfrac{1}{2}.$$

Now what about

$$P\left(6 \text{ appears } \textit{given} \text{ an even number appears}\right).$$

The above probability will be written

$$P(A|B) \;\; \text{to be read } P(A \text{ given } B).$$

How do we compute this probability? In fact we haven't defined it but we will compute it for this case and a few more cases using our intuitive sense of what it ought to be and then arrive at the definition later.

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

$$\left\{1, \boxed{2}, 3, \boxed{4}, 5, \boxed{6}\right\}$$

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

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

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

A very important example

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

$$P(\text{first heart}) = \frac{13}{52}.$$

Now, *what about the second heart*?

Many of you will come up with $\frac{12}{51}$ and

$$P(\heartsuit\heartsuit) = \left(\frac{13}{52}\right)\left(\frac{12}{51}\right).$$

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

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

What we really computed was the *conditional probability*

$$P\left(\heartsuit \text{ on } 2^{nd} \text{ deal} | \heartsuit \text{ on } 1^{st} \text{ deal}\right) = \frac{12}{51}.$$

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

$$P\left(\heartsuit \text{ on } 2^{nd} | \heartsuit \text{ on } 1^{st}\right) = \frac{12}{51}. \quad (1)$$

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

$$P(\heartsuit\heartsuit) = P\left(\heartsuit \text{ on } 1^{st}\right) P\left(\heartsuit \text{ on } 2^{nd} | \heartsuit \text{ on } 1^{st}\right) = \left(\frac{13}{52}\right)\left(\frac{12}{51}\right).$$

Two Basic Motivational Problems

I will use the next two problems to motivate the theory we will develop in the rest of this lecture. One way to solve these two problems will be contained in the HW Problem which follows this discussion. Another way is to use the general theory we are going to develop in the rest of this lecture.

1. What is
$$P\left(\heartsuit \text{ on } 1^{st} | \heartsuit \text{ on } 2^{nd}\right).$$
This is the reverse order of the conditional probability considered in Equation 1 where the answer was intuitively clear but for this reversed order it is not at all clear any more what the answer it.

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

An argument that works but I cannot justify

Here is a "proof" that will prove that this probability is the same as $P\left(\heartsuit \text{ on the } 1^{st}\right)$, namely $13/52 = 1/4$. First you know a card has been played so at first guess you might say $13/51$ *but* that card could have been a heart with probability $1/4$ so we subtract $1/4$ from the numerator to get $[13-(1/4)]/51 = \frac{51}{(4)(51)} = 1/4$ as I claimed. Intuitively there are 51 cards available and 12 and 3/4 hearts left. I know of no way to justify this rigorously - the sum $12+\frac{3}{4}$ amounts to adding a probability to a cardinality which I do not know how to justify. I think the sum is an expected value (to be defined later), the expected number of hearts left . However at the moment I do not see why the probability should be the ratio of this expected value divided by the number of cards available. Later in this Lecture and on the Homework problem that follows there will be rigorous proofs.

The Formal Mathematical Theory of Conditional Probability

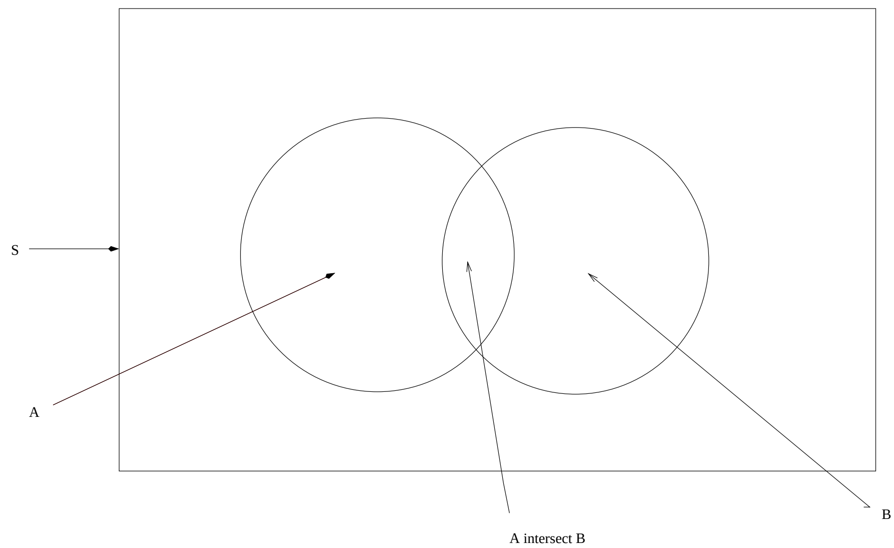

Figure 1: Conditional Probability Picture

$$\#(S) = n, \#(A) = a, \#(B) = b, \#(A \cap B) = c.$$

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

Problem
Compute $P(A|B)$. We are given B occurs so the conditional sample space is B. Once again we will compute a conditional probability without the formal mathematical definition of what it is.

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

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

We can rewrite this as

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

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.

The 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

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

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

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

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

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

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

Rewrite (**) as

$$P(A \cap B) = P(A)P(B|A) \tag{\#}$$

(#) is very important, more important than (**). It complements the formula

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

Now we know how P interacts with both of the basic binary operations $\cup$ and $\cap$. We will see that if A and B are *independent* then

$$P(A \cap B) = P(A)P(B).$$

I remember the above formula # as

$$P(\text{first } \cap \text{ second}) = P(\text{first})P(\text{second} \mid \text{first}).$$

Hopefully this will help you remember it.

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 $P(WRWB)$.

$$S = \{5W,\, 4B,\, 3R\}.$$

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

$$P(WRWB) = P(W) \cdot P(R|W) \cdot P(W|W \cap R) \cdot P(B|W \cap R \cap W).$$

Bayes' Theorem (pg.762)
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 often things." For example – we will get a second way (aside from the Problem at the end of this Lecture) to compute our favorite probability $P(\heartsuit \text{ on } 1^{st}|\heartsuit \text{ on } 2^{nd})$ because we know the "reversed probability" $P(\heartsuit \text{ on } 2^{nd}|\heartsuit \text{ on } 1^{st}) = 12/51$

First we will need a preliminary result.

The Law of Total Probability
Let $A_1, A_2, \cdots, A_k$, be mutually exclusive ($A_i \cap A_j = \emptyset, i \neq j$) and exhaustive $A_1 \cup A_2 \cup \cdots \cup A_k = S =$ the whole space). Then for any event B

$$P(B) = P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + \cdots + P(B|A_k)P(A_k). \qquad (\flat)$$

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

$$P(B) = P(B|A)(P(A) + P(B|A')P(A'). \qquad (\flat\flat)$$

Application to our ssecond basic problem

Now we can *prove*

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

Put

$$\begin{aligned} B &= \heartsuit \text{ on } 2^{nd} \\ A &= \heartsuit \text{ on } 1^{st} \\ A' &= \not\heartsuit \text{ on } 1^{st} \end{aligned}$$

Here we write $\not\heartsuit$ for nonheart. So

$$P(\not\heartsuit \text{ on } 1^{st}) = \frac{39}{52}$$
$$P(\heartsuit \text{ on } 2^{nd} | \not\heartsuit \text{ on } 1^{st}) = \frac{13}{51}.$$

Now

$$\begin{aligned} P(B) &= P(B|A)P(A) + P(B|A')P(A) \\ &= P(\heartsuit \text{ on } 2^{nd} | \heartsuit \text{ on } 1^{st}) P(\heartsuit \text{ on } 1^{st}) \\ &+ P(\heartsuit \text{ on } 2^{nd} | \not\heartsuit \text{ on } 1^{st}) P(\not\heartsuit \text{ on } 1^{st}) \\ &= \left(\frac{12}{51}\right)\left(\frac{13}{52}\right) + \left(\frac{13}{51}\right)\left(\frac{39}{52}\right) \end{aligned}$$

Add fractions:

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

Factor out 13

$$= \frac{(13)(12+39)}{(51)(52)} = \frac{(13)}{(52)} = \frac{(13)}{(52)} \qquad \textit{Done}!$$

($\flat$) is "easy" to prove but we won't do it. I did it in class for $k = 4$.

Now we can state Bayes' Theorem.

Bayes' Theorem
Let $A_1, A_2, \cdot, A_k$ be a collection of k mutually exclusive and exhaustive events. Then for any event B with $P(B) > 0$ we have

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

Again we won't give the proof for general k. For most applications $k = 2$ and in fact we don't the complicated formula for the denominator on the right-hand side (in fact it is just $P(B)$ which we can usually easily compute in applications).

Special Case $k = 2$
Suppose we have two events A and B with $P(B > 0)$. then

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

There is a sometimes better way to look at this formula. Namely

$$P(A|B) = \frac{P(A)}{P(B)}P(B|A).$$

The two equations are equivalent because of the Law of Total Probability Equation $\flat\flat$.

I will prove this case in the second form above because it shows what is really going on. The point is that $\cap$ is commutative so

$$A \cap B = B \cap A.$$

We will apply P to both sides of this equation and solve for $P(A|B)$ and thereby prove Bayes' Theorem. Indeed

$$\begin{aligned} &P(A \cap B) = \qquad P(B \cap A) \text{ so} \\ &P(A)P(B|A) = P(B)P(A|B) \text{ and solving for } P(A|B) \\ &P(A|B) = \frac{P(A)}{P(B)}P(B|A). \end{aligned}$$

Application to our first basic problem

We return to the computation of

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

We will use Bayes' Theorem. This is the obvious way to do it since we know the probability "the other way around" Put

$$A = (\heartsuit \text{ on } 1^{st}) \quad \text{and} \quad B = (\heartsuit \text{ on } 2^{nd}).$$

By Bayes' Theorem

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(\heartsuit\heartsuit)}{P(\heartsuit \text{ on } 2^{nd} \text{ with no other information})}$$

Now we know from earlier in this lecture that

$$P(\heartsuit\heartsuit) = \left(\frac{13}{52}\right)\left(\frac{12}{51}\right).$$

and we saw just before the proof of Bayes' Theorem (the solution of our first basic problem) that

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

Hence

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

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

This completes the solution of the second basic problem.

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

Problems

Problem 1

The point of this problem is to compute $P(\heartsuit on 1^{st}|\heartsuit on 2^{nd})$ and $P(\heartsuit on 1^{st})$ using an abstract approach counting pairs.

(i). Let S be the set of unordered pairs of distinct cards. S will be our sample space and the previous hard to understand problems will be solved by counting subsets of pairs with various conditions on one or both elements of the pair.

Compute $\#(S)$.

(ii). Let A be the subset of S consisting of pairs where both cards in the pair are hearts so

$$A = \{\heartsuit\heartsuit\} \subset S.$$

Compute

$$\#(A) \text{ and } P(A = \frac{\#(A)}{\#(S)}.$$

(iii) Let B be the subset of all pairs to that it the second card in the pair is a heart and the first card is anything. Compute $\#(B)$ and $P(B)$ solving 2. above (namely the probability that the second card is a heart with no condition on the first).

(iv) Compute $P\left(\heartsuit \text{ on } 1^{st}|\heartsuit \text{ on } 2^{nd}\right)$ by taking the ratio $P(A)/P(B)$, see the defining formula (*) two pages later, or else by computing within the conditional sample space S_B which is a subspace of the set of all pairs S.