1.4 Conditional Probability, the Multiplicative Law, the Law of Total Probability, and the Event Decomposition Method for Computing Probabilities

Conditional probability is a very important concept in probability theory and statistics because:

1) In practice, one often knows some partial information concerning the outcome of an experiment. Using that partial information, one can recalculate the probability. This is conditional probability.

2) Even if no partial information is known, conditional probabilities can be used to compute ordinary probabilities.

We begin with a motivational example where we can intuitively compute a conditional probability, see that it coincides with the actual definition, and recover it from two tables.

<u>Motivational Example</u> Consider two balanced 4-sided dice, one is blue, the other is red. The experiment is to roll them both and observe the two numbers they show.

$A :=$ the event that the sum is 5

$B :=$ the event that the blue die shows 2 or 3.

Determine the probability that the rolled sum is 5, given that the blue die shows 2 or 3, ie. find $P(A|B)$.

We encode the sample space as

$$S = \{(b,r) \mid b = \text{number on blue die}, r = \text{number on red die}\}$$
$$= \{(b,r) \mid b, r \in \{1,2,3,4\}\}$$

(1,4)	(2,4)	(3,4)	(4,4)
(1,3)	(2,3)	(3,3)	(4,3)
(1,2)	(2,2)	(3,2)	(4,2)
(1,1)	(2,1)	(3,1)	(4,1)

$A =$ event the sum is 5

$B =$ event the blue die shows a 2 or 3

$P(A|B)$ = probability the sum is 5 given that the blue die is 2 or 3

= "consider B as the new sample space, and find the proportion in B that has sum 5"

$= \dfrac{2}{8}$ ← Same!

Now notice that $\dfrac{P(A \cap B)}{P(B)} = \dfrac{2/16}{8/16} = \dfrac{2}{8}$

We can also encode the amounts and proportions in tables (like contingency tables), and see the conditional probability there.

Table of Amounts in Each Category:

	Blue 2 or 3	Blue not 2,3	Totals
Sum = 5	2	2	4
Sum ≠ 5	6	6	12
Totals	8	8	16

Table of Proportions in Each Category:

	Blue 2 or 3	Blue not 2,3	Totals
Sum = 5	2/16	2/16	4/16
Sum ≠ 5	6/16	6/16	12/16
Totals	8/16	8/16	16/16

The given event B that blue shows a 2 or 3 effectively means restrict attention to that column.

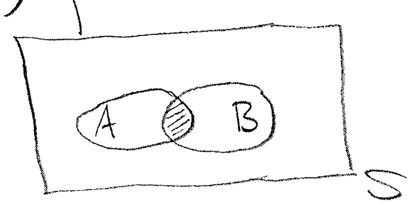
<u>Def</u> The <u>conditional probability of an event A given that event B has occurred</u> is defined by

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

provided $P(B) > 0$.

<u>Rem</u> We saw this in the motivational example.

<u>Exercise</u> If A and B are events in a finite sample space S, and each point is equally probable, prove that $P(A|B) = \dfrac{|A\cap B|}{|B|}$.



Another Motivational Example Consider the roll of a balanced 6-sided die. If a die roll is known to be odd, what is the probability it is a 1? Well, given that it is odd, it must be 1, 3, or 5. Thus the probability is $1/3$.

More formally: $A := \{1\}$

$B := \{1, 3, 5\}$

$$P(A|B) \overset{\text{def}}{=} \frac{P(A \cap B)}{P(B)} = \frac{1/6}{3/6} = \frac{1}{3} \checkmark$$

Ex A coin is flipped twice.

1) What is the probability that both flips are heads, given that the first flip is heads?

$S = \{HH, HT, TH, TT\}$

$A :=$ event that both flips are heads $= \{HH\}$

$B :=$ event that first flip is heads $= \{HH, HT\}$

$$P(A|B) \overset{\text{def}}{=} \frac{P(A \cap B)}{P(B)} = \frac{1/4}{2/4} = \frac{1}{2} \checkmark$$

2) What is the probability that both flips are heads given that <u>at least one</u> of the flips is heads?

S and A as above

$C :=$ event that at least one flip is heads

$= \{HH, HT, TH\}$

$$P(A|C) \overset{\text{def}}{=} \frac{P(A \cap C)}{P(C)} = \frac{1/4}{3/4} = \frac{1}{3}.$$

Intuitive Reason: Given that at least one flip is heads, TT is excluded, so there are only three remaining possible outcomes HH, HT, or TH.

Ex Discuss 2.77 a–c, h, i of WMS.

$A :=$ the offender has 10 or more years of education

$B :=$ the offender is convicted within two years of treatment.

$P(A) = .40$

$P(B) = .37$

$P(A \cap B) = .10$

$P(A|B) \overset{\text{def}}{=} \dfrac{P(A \cap B)}{P(B)} = \dfrac{.10}{.37} \approx .27$

$P(B|A) \overset{\text{def}}{=} \dfrac{P(B \cap A)}{P(A)} = \dfrac{.10}{.40} = .25$

$P(B|\bar{A}) \overset{\text{def}}{=} \dfrac{P(B \cap \bar{A})}{P(\bar{A})} = \dfrac{.27}{.60} = .45$

Thus, the probability of conviction within 2 years is much higher when the person has 9 or less years of education, than when the person has 10 or more years of education.

Note: In doing this problem, notice that the given condition distinguished a column or row for us to focus attention on.

Rem From the above example about drugs, we see that $P(A|B)$ and $P(B|A)$ are not necessarily equal.

Conditional probability is not symmetric.

Proposition If $P(B)>0$, then $P(-|B)$ is a probability measure. That is,

1) for each event A, $P(A|B) \geq 0$
2) $P(S|A)=1$
3) for any pairwise mutually exclusive events $A_1, A_2, A_3, \ldots$ we have
$$P\left(\bigcup_{i=1}^{\infty} A_i \,\Big|\, B\right) = \sum_{i=1}^{\infty} P(A_i|B).$$

Proof in Book

Rem Since $P(-|B)$ is a probability measure, it has all the consequent properties. For instance,
$$P(\bar{A}|B) = 1 - P(A|B)$$
and
$$P(C \cup D|B) = P(C|B) + P(D|B) - P(C \cap D|B).$$

<u>Thm</u> (Multiplicative Law) The probability that events A and B in a sample space both occur is

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

which is also equal to

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

<u>Ex</u> Anitra is deciding between a topology course and an algebra course. Her probability of earning an A in topology is $2/3$, while her probability of earning an A in algebra is $1/2$. Anitra can't make up her mind between these two fascinating courses, so she bases her decision on a coin flip.

What is the probability she gets an A in topology?

$A :=$ event she gets an A in whatever course she takes

$B :=$ event she takes topology

$$P(A \cap B) \overset{\text{Thm}}{=} P(B) \cdot P(A|B) = \frac{1}{2} \cdot \frac{2}{3} = \frac{1}{3}$$

(Notice we could <u>not</u> have used the first half of the Multiplication Law

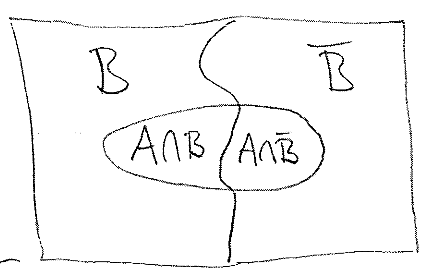
$$P(A\cap B)=P(A)P(B|A)$$

because we don't know $P(A)$, nor $P(B|A)$).

<u>Thm</u> (Law of Total Probability) Suppose A and B are events in a sample space. Then

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

<u>Pr</u>



S

First note $S=B\cup\overline{B}$ and $B\cap\overline{B}=\emptyset$ (so $B,\overline{B}$ form a <u>partition</u> of S).

Then $A=A\cap S=A\cap(B\cup\overline{B})$

$$=\underbrace{(A\cap B)\cup(A\cap\overline{B})}_{\text{disjoint because}}$$

disjoint because

$\emptyset\subseteq(A\cap B)\cap(A\cap\overline{B})\subseteq B\cap\overline{B}=\emptyset.$

Thus

$$
\begin{aligned}
P(A) &= P((A \cap B) \cup (A \cap \bar{B})) \\
&= P(A \cap B) + P(A \cap \bar{B}) \\
&= P(B)P(A|B) + P(\bar{B})P(A|\bar{B}) \\
&= P(A|B)P(B) + P(A|\bar{B})P(\bar{B}) \checkmark
\end{aligned}
$$

$\square$

Ex WMS 2.110 First we define the events whose probabilities are given, and the event whose probability we want to find.

The experiment is to pick an item from the day's production, and observe 1) which line it is from
2) whether defective or not defective

$L_1 :=$ event that the item was produced on line 1

$L_2 :=$ event that the item was produced on line 2

$N :=$ event that the item is not defective

$D :=$ event that the item is defective.

Want $P(N)$.

$$P(N) \overset{\text{Thm}}{=} P(N \mid L_1)P(L_1) + P(N \mid L_2)P(L_2)$$

Law of Total Probability

$$= (1 - P(\bar{N} \mid L_1))P(L_1) + (1 - P(\bar{N} \mid L_2))P(L_2)$$

Since $P(- \mid L_1)$ and $P(- \mid L_2)$ are probability measures, the Complement Law holds for them.

$$= (1 - .08).4 + (1 - .10).6$$

$$= .908$$

$\Longrightarrow$ The probability a randomly selected item will not be defective is .908. Done.

Q. What actually was the sample space and the events?

$S = \{1d, 1nd, 2d, 2nd\}$

$L_1 = \{1d, 1nd\}$ $\quad$ $L_2 = \{2d, 2nd\}$

$N = \{1nd, 2nd\}$ $\quad$ $D = \{1d, 2d\}$

Rem Notize that we did not have to find the probabilities of the individual sample points to apply the Law of Total Probability.

Ex WMS 2.111 First we define the events whose probabilities are given, and the event whose probability we want to find.

The experiment is to randomly select a potential customer and observe if she/he purchases the product.

M := set of people who see the magazine ad

T := set of people who see the TV ad

B := set of people who buy the product

$M \cup T$ = set of people who see the magazin ad, the TV ad, or both

Want $P(B)$.

$$P(M \cup T) = P(M) + P(T) - P(M \cap T) = 1/50 + 1/5 - 1/100 = .21$$

$$P(B) = \underbrace{P(B \mid M \cup T)}_{1/3} \underbrace{P(M \cup T)}_{.21} + \underbrace{P(B \mid \overline{M \cup T})}_{1/10} \underbrace{P(\overline{M \cup T})}_{1-.21}$$

$$= \frac{1}{3} \cdot (.21) + \frac{1}{10} \cdot (1 - .21)$$

$= .149$ $\Rightarrow$ The probability that a randomly selected potential buyer will buy the product is .149.

Methods for Computing the Probability of an Event: the Sample Point Method and the Event Decomposition Method

If we have an event $A \subseteq S$, how do we compute $P(A)$?

Sample Point Method: 1) Determine the sample space.

2) Determine the probability of each individual sample point, i.e. find $P(\{a\})$ for all $a \in A$.

3) Then $P(A) =$ sum of the probabilities of the sample points in A

Ex When the sample space S is finite, and each sample point is equally likely, we have seen $P(A) = \dfrac{|A|}{|S|}$. For instance, for the role of a fair six-sided die, $P(\text{roll is odd}) = P(\{1,3,5\})$

$$= \frac{|\{1,3,5\}|}{|\{1,2,3,4,5,6\}|} = \frac{3}{6} = \frac{1}{2}$$

Ex Now consider the roll of die that is biased toward even numbers, especially 4. Suppose the probabilities of the outcomes are

outcome	probability
1	.1
2	.2
3	.1
4	.3
5	.1
6	.2

Then
$$\begin{aligned} P(\text{roll is odd}) &= P(\{1,3,5\}) \\ &= P(\{1\}) + P(\{3\}) + P(\{5\}) \\ &= .1 + .1 + .1 \\ &= .3 \quad \text{which is } \underline{\text{not}} \ \frac{|A|}{|S|} \end{aligned}$$

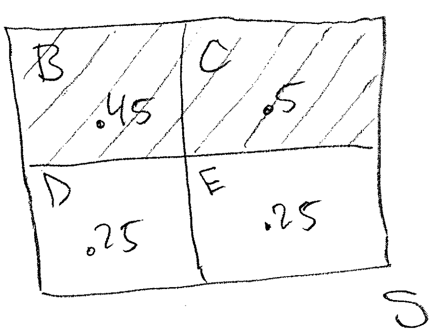
$$\begin{aligned} P(\text{roll is } < 5) = P(\{1,2,3,4\}) &= P(\{1\}) + P(\{2\}) + P(\{3\}) + P(\{4\}) \\ &= .1 + .2 + .1 + .3 \\ &= .7 \quad \text{which is } \underline{\text{not}} \ \frac{|A|}{|S|}. \end{aligned}$$

Another way to compute $P(A)$ is the following, which does not require us to know the probability of the individual sample points.

<u>Event Decomposition Method</u>: 1) Express A in terms of unions, intersections, and/or complements of events whose probabilities or conditional probabilities are known.

2) Use the Inclusion-Exclusion Principle, Complement Law, Multiplicative Law, or Law of Total Probability to compute $P(A)$ via the expression in step 1.

<u>Ex</u> Suppose A is decomposed as $A = B \cup C$ and the following probabilities are known.



Then
$$
\begin{aligned}
P(A) &= P(B \cup C) \\
&= P(B) + P(C) - P(B \cap C) \\
&= .45 + .5 - \underbrace{P(\emptyset)}_{=0} \\
&= .5
\end{aligned}
$$