

Chapter 10 : Probability

Probability of an event A is interpreted as the *long run* proportion of times that event A occurs. See the text, and in particular example 10.2 about fair coin tossing.

As a second example of this type for the lecture class we consider the problem of rolling a fair die. A die is a six sided cube with faces with dots on them, one dot on one side, two dots on a second side, three dots on a third side, etcetera up to six dots on the sixth side.

Each of the six sides has the same chance of occurring. Thus each side has a chance $\frac{1}{6} = .1666\dots = .167$ (rounded to 3 decimal places). Figure 1 shows the running means or averages of $N = 6000$ die rolls. Table 1 gives the first 20 die rolls and the running means or running averages of the numbers of one observed, and of the numbers of twos observed.

After 2 die rolls there are no ones and 1 two observed, so these proportions are $\frac{0}{2}$ and $\frac{1}{2}$ respectively. On die roll 10 we observe a one, and so the observed proportions of ones and twos are now $\frac{1}{10}$ and $\frac{1}{10}$ respectively. On die roll 11 we observe a two, and so the observed proportions of ones and twos are now $\frac{1}{11}$ and $\frac{2}{11}$ respectively. Figure 1 shows these for the $N = 6000$ die rolls. The interesting thing in the picture is that the observed sample proportions of the event “1” and event “2” settle down to near the value of $\frac{1}{6}$.

A similar experiment and running sample proportion of the event “head” will also settle down to a value of near $\frac{1}{2}$, as shown in the text.

This idea and *interpretation of probability* is called the frequentist interpretation, that is it refers to long run relative proportions. It applies to classic gambling games and lottery games, as well as any experiment where one can envision repetitions of the experiment or game.

roll number	outcome	proportion of 1	proportion of 2
1	3	0.000	0.000
2	2	0.000	0.500
3	5	0.000	0.333
4	6	0.000	0.250
5	6	0.000	0.200
6	4	0.000	0.167
7	3	0.000	0.143
8	6	0.000	0.125
9	5	0.000	0.111
10	1	0.100	0.100
11	2	0.091	0.182
12	4	0.083	0.167
13	2	0.077	0.231
14	3	0.071	0.214
15	6	0.067	0.200
16	5	0.062	0.188
17	1	0.118	0.176
18	2	0.111	0.222
19	3	0.105	0.211
20	1	0.150	0.200

Table 1: Table Measurement from First Class

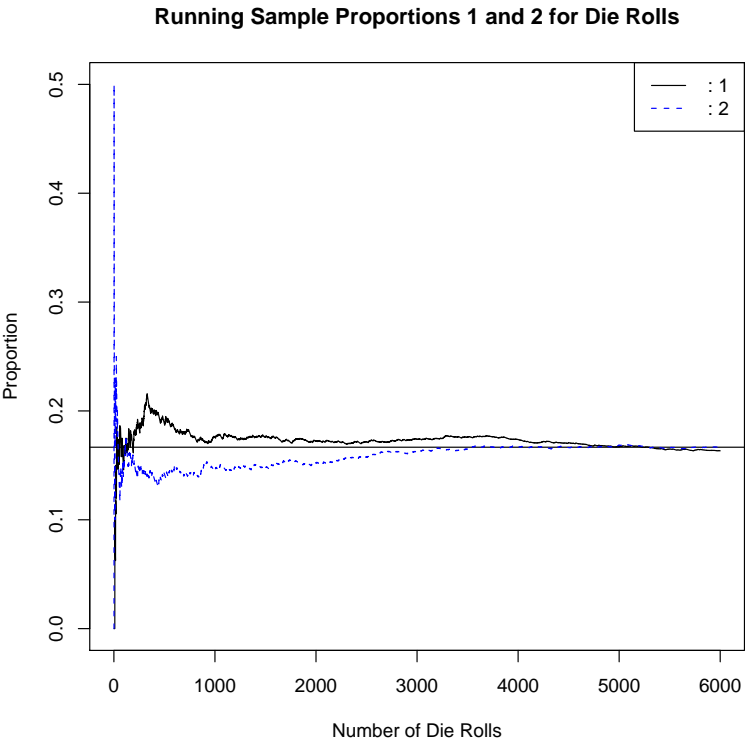


Figure 1: Running Means for 6000 Fair Rolls of a Die

We call a phenomenon *random* if the outcomes are uncertain but there is a regular distribution of outcomes in large numbers of repetitions.

While the outcome of a coin toss or of die rolling is *random*, that is not predictable with certainty, they are predictable in a distributional sense. That is we can predict the probability that the outcome will fall into a certain set with a given probability. For example when we roll a fair die, the probability that the outcome will be “4” is $\frac{1}{6}$.

For example in Lotto 6/49, there are 49 balls, and 6 are chosen to determine a winning ticket. There are 13,983,816 such combinations of 6 different numbers out of the 49, and the chance of any one particular number being chosen is $\frac{1}{13983816} = 0.0000000715112$. Sometimes this is stated as odds of 13,983,816 to 1. By way of contrast the chance of being struck by lightning in a given year is $\frac{1}{700,000} = 0.00000142857$ (as given in the URL <http://www.lightningsafety.noaa.gov/medical.htm> for government statistics in the USA for the year 2000).

This says that a person is 20 times more likely to be struck by lightning in a given year, than winning with a single ticket in Lotto 6/49.

Probability Models

Many gambling games are based on coins, cards or dice. These games have only finitely many outcomes, that is one can list all the outcomes.

Probability model

- sample space S = set of all possible outcomes
- event = outcome or set of outcomes of the random phenomenon.

An event is a subset of the sample space S

- Probability model is a mathematical description of the random phenomenon consisting of two parts : (i) sample space and (ii) a rule assigning probabilities to events

The rules of probability are mathematical rules and applies no matter what interpretation of probability might be. So far we have only considered the most useful, that is the frequentist interpretation of probability.

Consider a game of rolling two fair dice. These outcomes are represented in Table 2. There are 36 possible pairs for the two dice. Since the dice are fair each of the 36 outcomes has the same chance of occurring, so each has a chance $\frac{1}{36}$ of occurring or happening.

Table 2: Outcomes for rolling two dice

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

In some games one only cares about the total on the two faces. In this case we can calculate the outcomes on the possible rolls of the dice.

	die number 2					
die number one \	1	2	3	4	5	6
1	1+1=2	1+2=3	1+3=4	1+4=5	1+5=6	1+6=7
2	2+1=3	2+2=4	2+3=5	2+4=6	2+5=7	2+6=8
3	3+1=4	3+2=5	3+3=6	3+4=7	3+5=8	3+6=9
4	4+1=5	4+2=6	4+3=7	4+4=8	4+5=9	4+6=10
5	5+1=6	5+2=7	5+3=8	5+4=9	5+5=10	5+6=11
6	6+1=7	6+2=8	6+3=9	6+4=10	6+5=11	6+6=12

From this we see there are only outcomes for the sum of the two dice, along with their probabilities as

total spots	2	3	4	5	6	7	8	9	10	11	12
Probability	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

Table 3: Distribution for sum of two fair die rolls

Probability Rules

- Rule 1 : The probability $P(A)$ is between 0 and 1, that is $0 \leq P(A) \leq 1$
- Rule 2 : The sample space S has probability 1, that is $P(S) = 1$.
- Rule 3 : Two events A and B are disjoint if they have no outcomes (elements) in common. If two events A and B are disjoint the

$$P(A \text{ or } B) = P(A) + P(B)$$

- Rule 4 : For any event A

$$P(A \text{ does not occur}) = 1 - P(A)$$

Aside : For some who have taken appropriate course you may have used the notation \emptyset = the empty set, the symbol \cup = union, and \cap = intersection. In this notation we say A and B are disjoint if

$$A \cap B = \emptyset$$

We also use

$$A \text{ or } B = A \cup B$$

so that Rule 3 says

Rule 3 : IF events A and B are disjoint then

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

End of Aside

How do we use these Rules to understand our dice game calculations above?

If A_3 represents the event of rolling the sum on the two die faces totalling 3, then A_3 represents the outcome of the game being either

- die 1 comes up 1 and die 2 comes up 2 OR
- die 1 comes up 2 and die 1 comes up 1

These two elementary outcomes are disjoint, that is if the first happens the second cannot happen, and vice versa. Thus using Rule 2 we calculate

$$\begin{aligned}
 P(\text{sum of die faces is 3}) &= P(\text{die 1 comes up 1 and die 2 comes up 2}) \\
 &\quad + P(\text{die 1 comes up 2 and die 2 comes up 1}) \\
 &= \frac{1}{36} + \frac{1}{36} \\
 &= \frac{2}{36}
 \end{aligned}$$

Using the same Rule we calculate

$$\begin{aligned}
 P(\text{sum of die faces is 4}) &= P(\text{die 1 comes up 1 and die 2 comes up 3}) \\
 &\quad + P(\text{die 1 comes up 2 and die 2 comes up 2}) \\
 &\quad + P(\text{die 1 comes up 3 and die 2 comes up 1}) \\
 &= \frac{1}{36} + \frac{1}{36} + \frac{1}{36} \\
 &= \frac{3}{36}
 \end{aligned}$$

An easier to use notation is to consider a random variable with a given distribution.

Suppose X represents the outcome for the total on the faces of the two dice, and suppose the two dice are fair.

Notation We usually use capital letters from near the end of the alphabet to name random variables. These are typically X, Y, Z and when we need extra letters T, U, V, W . Sometimes we use a name that might be more helpful such as T when our random variable involves time. We use Z when talking about a standard normal random variable (see Chapter 3 and Table A).

For rolling a pair of dice then we write

$$P(\text{sum of die faces is } 4) = P(X = 4)$$

and more generally

$$P(\text{sum of die faces is } x) = P(X = x)$$

for the various possible values of x , in this case $x = 2, 3, 4, \dots, 10, 11, 12$. Capital X represents the name of the random variable and lower case (little) x represents some real number. In the context it should be clear.

What is the probability that you roll a 7 or 11? It is

$$\begin{aligned} P(X = 7 \text{ or } X = 11) &= P(X = 7) + P(X = 11) \\ &= \frac{6}{36} + \frac{2}{36} \\ &= \frac{8}{36} \end{aligned}$$

where we use the information from Table 3.

The text discusses Benford's Law. More information on this can be found at the URL

<http://www.intuitor.com/statistics/Benford%27s%20Law.html>

This example is interesting in that real data, such as accounting records, has a particular distribution for the first non-zero digit. Data that does not follow this is suspicious and suggests that the data is made up. A forensic audit team may record a large sample of first digits and compare the observed distribution (sample proportions) of these to the distribution predicted by Benford's Law. If the observations are inconsistent with this Law then there is evidence that the data is unusual. Further investigations can then be made.

Continuous Probability Models

A *continuous probability model* assigns probability for data observed in an interval as the area under a *probability density curve or function* for this interval. We have seen one such example, the normal curves in Chapter 3. Figure 2 shows the area under the standard normal curve for the interval $[-1, 1.5]$. This area can be calculated from Table A, where we have to look for the z values of -1 and 1.5. The area is then

$$P(Z \leq 1.5) - P(Z \leq -1) = 0.933 - 0.159 = 0.774$$

We interpret this as the probability that a standard normal random variable falls into the interval $[-1, 1.5]$.

If X has a normal distribution, mean $= \mu = 70$, and standard deviation $= \sigma = 8$, then

$$\begin{aligned} P(62 \leq X \leq 82) &= P\left(\frac{62 - 70}{8} \leq \frac{X - 70}{8} \leq \frac{82 - 70}{8}\right) \\ &= P(-1 \leq Z \leq 1.5) \\ &= 0.933 - 0.159 = 0.774 \end{aligned}$$

where in the second last line the random variable Z has a standard normal distribution (recall this property of the relation of a general normal distribution to the standard normal distribution).

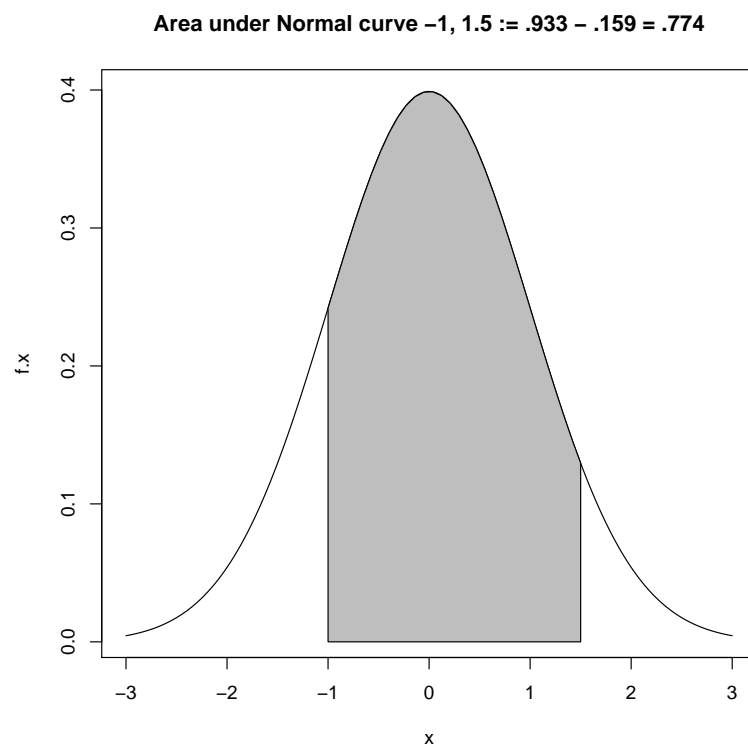


Figure 2: Area under normal curve for interval $[-1, 1.5]$

Another very useful probability curve is the uniform density. Figure 3 shows this curve and the area under it for the interval $[.5, .7]$. The Uniform 0, 1 curve is of constant height 1 over the interval 0 to 1, and zero everywhere else. If X represents the uniform(0,1) random variable then

$$P(.5 \leq X \leq .7) = .7 - .5 = .2$$

For more general numbers a and b where $0 \leq a < b \leq 1$ we have

$$P(a \leq X \leq b) = b - a .$$

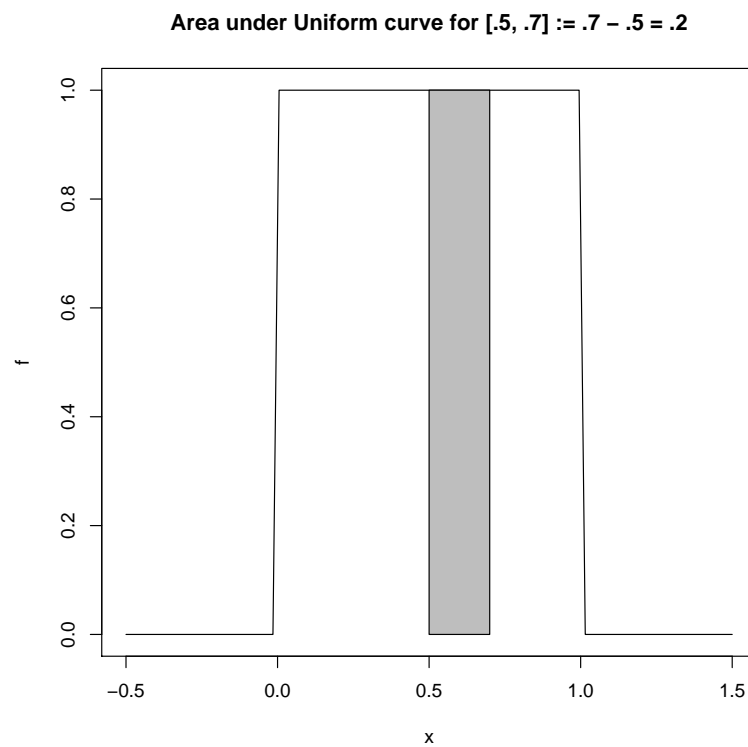


Figure 3: Area under normal curve for interval $[.5, .7]$

Random Variables

A *random variable* is a variable whose value is a numerical outcome of a random phenomenon.

A *probability distribution* of a random variable X is a rule that tells us what values X can take and how to assign or calculate probabilities that X takes on specific values or that X falls into specific intervals.

This is what we have done for some calculations above, in terms of normal random variables, and earlier for some dice games, for example the distribution of the sum or total on two die faces given in Table 3.

We use a symbol or name such as capital X as the *name* of the random variable. The distribution calculates probabilities for events that are given in terms of the random variables. Different random variables generally need different names, depending on the context. A non mathematical analogy is that a person or individual can have a name, say Aragorn (who is the king who returns in the *Lord of the Rings* and is the most noble human in that story). Other individuals in the story are Frodo, Sam, Gandalf, and Bill the Pony, amongst others. The name of the individual is useful for identifying the individual, say Aragorn, but it does not give properties (nobility) or numerical characteristics (age, height, weight) of this individual.

There is one other common interpretation of probability, the so called *personal probability*. This is helpful when one is considering a one time event, and not a repetition of many trials. For example you might ask what is the probability that I will get an A in this course or that I will pass this course. *Personal probability* expresses your judgement on how likely an outcome is.

It is often thought of as to what you would be willing to bet on the outcome to the point where you feel it is not to your advantage of taking a bet. One branch of philosophy is concerned with these ideas, and we not delve further into this in our course.

Personal probability obeys the same mathematical rules as probability under any other interpretation, such as the frequentist interpretation.

Chapter 12 : General Rules of Probability

This material deals with

- Independence and multiplication rule
- General addition rule
- Conditional probability
- General multiplication rule

which is only the first part of Chapter 12.

Recall from earlier in Chapter 10 the Probability Rules

- Rule 1 : The probability $P(A)$ is between 0 and 1, that is $0 \leq P(A) \leq 1$
- Rule 2 : The sample space S has probability 1, that is $P(S) = 1$.
- Rule 3 : Two events A and B are disjoint if they have no outcomes (elements) in common. If two events A and B are disjoint the

$$P(A \text{ or } B) = P(A) + P(B)$$

- Rule 4 : For any event A

$$P(A \text{ does not occur}) = 1 - P(A)$$

Figures 4 and 5 illustrate the notion of intersecting and disjoint sets in a Venn diagram. Rule 3 says if the sets or events are disjoint then probabilities add.

Return to the die rolling game where two fair dice are rolled and consider getting a total of 3 on the two dice. This can happen if either A or B happens, where

A = roll 1 on first die and 2 on second die

B = roll 2 on first die and 1 on second die

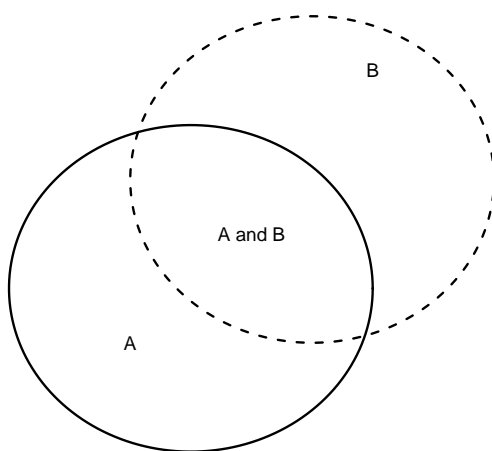
Venn Diagram : Intersecting sets

Figure 4: Venn Diagram Intersecting Sets

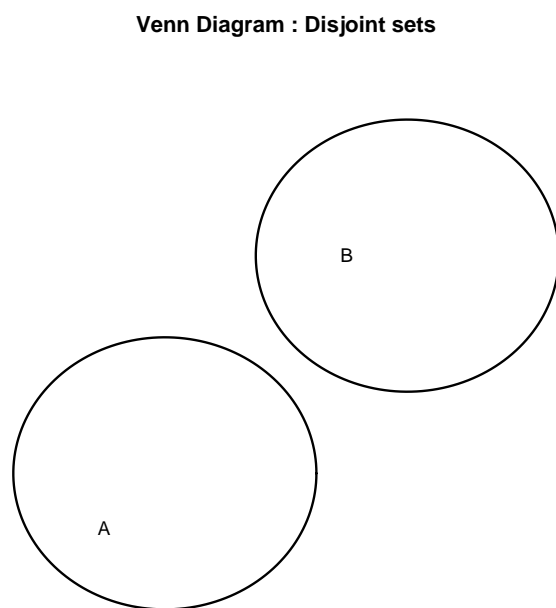


Figure 5: Venn Diagram Disjoint Sets

Notice that in this game if A happens then B cannot happen, and if B happens then A cannot happen. In this case A and B are disjoint, that is they have no outcomes (of the pair of die rolls) in common.

Thus

$$\begin{aligned} P(\text{roll a sum of 3}) &= P(A \text{ or } B) \\ &= P(A) + P(B) \\ &= \frac{1}{36} + \frac{1}{36} = \frac{2}{36} \end{aligned}$$

Underlying this is a notion of *independence*, sometimes called statistical independence. Consider events

$$\begin{aligned} E_1 &= \text{roll 1 on first die} \\ E_2 &= \text{roll 2 on second die} \end{aligned}$$

When the two die are rolled, the outcome on the first die does not influence the outcome on the second die. Thus we interpret

$$\begin{aligned} &P(\text{roll 1 on first die and 2 on second die}) \\ &= P(E_1 \text{ happens and } E_2 \text{ happens}) \\ &= P(E_1 \text{ happens}) * P(E_2 \text{ happens}) \\ &= P(E_1) * P(E_2) \\ &= \frac{1}{6} * \frac{1}{6} = \frac{1}{36} \end{aligned}$$

This is a general rule.

Multiplication Rule for Independent Events : Events A and B are independent if and only if

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

One way of seeing how this makes sense is as is to think about the die rolling game again. Out of all possible first die rolls, a fraction $\frac{1}{6}$ comes up as “1”. Of these $\frac{1}{6}$ of all possible outcomes, only a fraction $\frac{1}{6}$ of these, that is $\frac{1}{6}$ of the $\frac{1}{6}$ of the outcomes for rolling two dice has a “1” on the first die and a “2” on the second die.

	Domestic (N.A.)	Import	Total
light truck	0.47	0.07	0.54
car	0.33	0.13	0.46
Total	0.80	0.20	1.00

Table 4: Import versus North American (domestic) vehicles)

General addition rule for intersecting sets; recall Figure 4.

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$$

To help us make sense of this, in Figure 4, if think of counting areas,

- the parts of A get counted once where it does not intersect B , once for where it intersects B
- the parts of B get counted once where it does not intersect A , once for where it intersects A

Thus the part where A and B intersect (which is the same as where B and A intersect) gets counted twice, whereas the other parts of A and B get counted once, so we need to subtract off one of the extra countings of A intersect B .

To help us with this we use an example from the text, Example 12.5 in Edition 4.

Origin of light trucks and cars sold are given in Table 4. In this Table, domestic (N.A.) means domestic interpreted as North America (N.A.). The entries in the main table give the proportion (probability) of the type and origin of vehicle. The *margins* give the marginal probabilities, the same idea as the *marginal totals* in our earlier discussion of two way tables.

From this table we can now calculate for example the probability that a vehicle sold is either domestic or a light truck. The AND is given in capital letters just to emphasize

this.

$$\begin{aligned}
 &P(\text{ domestic or light truck}) \\
 &= P(\text{ domestic }) * P(\text{ light truck }) - P(\text{domestic AND light truck}) \\
 &= 0.80 + 0.54 - 0.47 \\
 &= 0.87
 \end{aligned}$$

As we studied with the two way tables in Chapter 6, we also need to consider a notion of *Conditional Probability*.

Conditional Probability : When $P(A) > 0$ the conditional probability of B given that the event A happens is

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

If $P(A) = 0$, that is event A has zero chance of occurring (impossible to occur) then we would never be interested in calculating the conditional probability of B given that A occurs.

Using Table 4 we can ask what is the conditional probability that a vehicle is imported, given that it is a car. In terms of our thinking of this in our two way table discussion this would be the relative proportion of *imported cars* amongst only the *cars*. Except for the scaling this is

$$\frac{0.13}{0.33 + 0.13} = \frac{0.13}{0.46} = 0.283 .$$

This is also given by the conditional probability formula

$$\begin{aligned}
 &P(\text{ imported} | \text{ car}) \\
 &= \frac{P(\text{ imported and car})}{P(\text{ car })} \\
 &= \frac{0.13}{0.46} = 0.283 .
 \end{aligned}$$

Finally using the formula for conditional probability, multiplying both sides by $P(A)$ we have

The General Multiplication Rule

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

Note there is also a conditional probability formula for the probability of A given B so we also have

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

and

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

which is the same thing just interchanging the role of A and B .

This rule is often useful for calculating probabilities for two or multistage games.

Consider the game of choosing marbles, without replacement from a box. At the beginning, stage 0, the box contains 4 white marbles and 3 red marbles. At this stage we know the probabilities of the colour of the first ball chosen or drawn

$$\begin{aligned} P(\text{ white marble on stage 0}) &= \frac{4}{7} \\ P(\text{ red marble on stage 0}) &= \frac{3}{7} \end{aligned}$$

These are also illustrated in Figure 6.

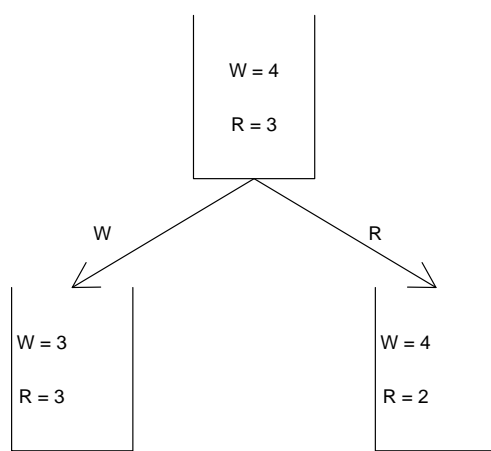


Figure 6: Choosing Marbles Without Replacement

Since the physical content of the box changes at the next stage, stage 1, it turns out to be easier to calculate the conditional probabilities of the colour of the second marble chosen (from the box at this corresponding stage of the game) conditional on the colour of the marble chosen at the first stage.

$$\begin{aligned}
 P(\text{ white marble on stage 1} \mid \text{ white on stage 0}) &= \frac{3}{6} \\
 P(\text{ white marble on stage 1} \mid \text{ red on stage 0}) &= \frac{4}{6} \\
 P(\text{ red marble on stage 1} \mid \text{ white on stage 0}) &= \frac{3}{6} \\
 P(\text{ red marble on stage 1} \mid \text{ red on stage 0}) &= \frac{2}{6}
 \end{aligned}$$

We can then calculate the probability of two whites, that is white on first draw (from the box at stage 0) and white on second draw (the box at stage 1)

$$\begin{aligned}
 &P(\text{ white on first draw and white second draw}) \\
 &= P(\text{ white marble on stage 1} \mid \text{ white on stage 0}) * P(\text{ white on draw 1}) \\
 &= \frac{3}{6} * \frac{4}{6} \\
 &= \frac{2}{3} = 0.667 .
 \end{aligned}$$

There are other important uses for conditional probability, but they are not discussed in this introductory course.