Chapter 6

Conditional Probability

6.1 Introduction

The concept of conditional probability is a fascinating one. What if one noticed that an event occurred and we wish to follow that with another event. Does the probability of the second event get influenced or affected by the first event? Furthermore, suppose there is more than one descriptor for an experiment– does one descriptor affect or have some bearing on another?

A couple of examples might be best to illuminate the idea:

Example 6.1.1. *You have a fair standard 52-card bridge deck of cards. The descriptors on each card are rank and suit.*
A. Draw a card. The probability that card is a jack is $\frac{4}{52}$.
B. Draw a card. Assume you know the card is a spade. The probability that spade card is a jack is $\frac{1}{13}$ *(the nomenclature is the probability the card is a jack given the card is a spade).*
C. Draw a card. Assume you know the card is a jack. The probability that jack is a spade is $\frac{1}{4}$.
D. Draw a card. Assume you know the card is a jack. The probability that jack is a 5 is $\frac{0}{4}$.

Notice how the probabilities were not constant. The example just presented was such that the descriptors were temporally concurrent. What if the experiment is such that we have temporally distinct events? How does that work? Another illustration might help.

Example 6.1.2. *You have an urn with 5 red balls, 2 blue balls, and 3 white balls.*
A. Draw a ball from the urn. The probability that ball is red is $\frac{5}{10}$.
B. Draw a ball from the urn; then another. Suppose the first ball was red. The probability that the second ball is red is $\frac{4}{9}$ *(the nomenclature is the probability the second ball is red given the first ball is red).*
C. Draw a ball from the urn; then another. Suppose the first ball was red. The probability that the second ball is blue is $\frac{2}{9}$.
D. Draw a ball from the urn; note the colour. Put it back and shake up the urn. Now pick a ball (another). Suppose the first ball was red. The probability that the second ball is blue is $\frac{2}{10}$.

The examples suggest the definitions and concepts that follow.

6.2 Definitions and Concepts

Definition 6.2.1. Let S be a well defined sample space; E an event. E is a non-trivial event means $Pr(E) \in (0,1)$.

Definition 6.2.2. Let S be a well defined sample space; E and F be events where $Pr(F) = 0$. The probability of E given F has occurred is 0. The notation for it is $Pr(E|F)$.

Definition 6.2.3. Let S be a well defined sample space; E and F be events where $Pr(F) \neq 0$. The probability of E given F has occurred is $\dfrac{Pr(E \cap F)}{Pr(F)}$. The notation for it is $Pr(E|F)$.

Theorem 6.2.1. *Let S be a well defined sample space; and E a non-trivial event. Therefore it is the case that E^c is a non-trivial event.*

There are ways to draw diagrammes, graphs, or charts well so that the graph, chart, or diagramme[1] illustrates the probability in such a manner as to enhance both one's understanding of the concept and portray accurately the scenario.

Example 6.2.1. *$U = \mathbb{R}$ and $S = [0, 10]$, whilst $E = (0,1), F = [1,4), J = (1,5], \ldots$*
Draw the set of real numbers ($\mathbb{R}$). The points on the line. Proceed.

Example 6.2.2. *You roll a pair of fair dice and view the upward faces of the dice such that an event is defined as the sum of the values of the two dice faces. The probability of rolling a five (call it E) on one of the dice given you rolled a sum that is prime (call it M) is illustrated in the table.*

first die →	*1*	*2*	*3*	*4*	*5*	*6*
second die ↓						
1	*2*	*3*	*4*	*5*	*6*	*7*
2	*3*	*4*	*5*	*6*	*7*	*8*
3	*4*	*5*	*6*	*7*	*8*	*9*
4	*5*	*6*	*7*	*8*	*9*	*10*
5	*6*	*7*	*8*	*9*	*10*	*11*
6	*7*	*8*	*9*	*10*	*11*	*12*

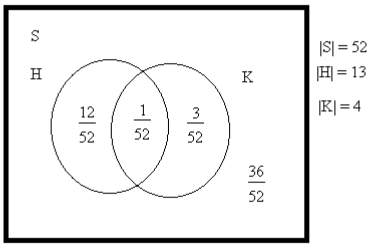
There are 36 entries

eleven of which is where a five is rolled on one of the two dice ($Pr(E) = \frac{11}{36}$);
fifteen are where a sum is rolled that is prime ($Pr(M) = \frac{15}{36}$); and, four where a five is rolled on a die and a prime sum was rolled ($Pr(E \cap M) = \frac{4}{36}$). So,

$$(Pr(E|M) = \frac{Pr(E \cap M)}{Pr(M)} = \frac{4}{36} \div \frac{15}{36} = \frac{4}{15})$$

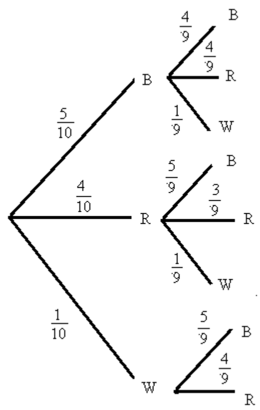
[1]Recall from Math 224 we studied Hasse diagrammes and Venn diagrammes. Other types of diagrammes of use are Tree diagrammes and Euler diagrammes. It is the case that Hasse diagrammes are useful for partial orders and Euler diagrammes for syllogistic logic; but not for probability. For probability we use either Venn diagrammes or Tree diagrammes.

Example 6.2.3. *You draw a card from a fair bridge deck of cards. The probability of drawing a king (call it K) given you drew a heart (call it H) is illustrated in the Venn Diagramme.*



Note $Pr(K) = \frac{4}{52}$*;* $Pr(H) = \frac{13}{52}$*; and,* $Pr(K \cap H) = \frac{1}{52}$.
So, $Pr(K|H) = \frac{1}{4}$ *which can be seen in the Venn diagramme or through the algebra of* $Pr(K|H) = \frac{Pr(K \cap H)}{Pr(H)} = \frac{1}{52} \div \frac{4}{52} = 0.25$

Example 6.2.4. *There is an urn with 5 blue, 4 red, and a white ball. You draw a ball from the urn and note the colour. You draw a second ball from the urn and note its colour. The probability of drawing a red ball (call it R) second given you drew a blue ball first (call it B) is illustrated in the Tree Diagramme.*



Note $Pr(R_{second}|B_{first}) = \frac{4}{9}$ *which can be seen in the Tree diagramme as the second branch labeled R off of the top primary branch (B).*

Notice that the illustrations sometimes have the conditional probabilities labeled within and other times the conditional probability must be derived after the illustration. Nonetheless, for computational problems the diagrammes are, I opine, most helpful. For the theoretical, I am not sure.

6.3 Claims & Exercises

So, consider the following claims to prove or disprove.
All claims that follow are $\mathbb{B}$.

Claim 6.3.1. Let S be a well defined sample space; E, F, and G are events; $E \subseteq F$; and, E and G are mutually exclusive. Therefore it is the case that
$Pr(E \cup F \cup G) = Pr(F) + Pr(G) - Pr(F \cap G)$.

Claim 6.3.2. Let $U = \mathbb{R}$ and $S \subseteq U$ be a well defined sample space with $A \wedge B$ events. Therefore it is the case that $0 \leqslant Pr(B|A) \leqslant 1$.

Claim 6.3.3. Let $U = \mathbb{R}$ and $S \subseteq U$ be a well defined sample space with $A \wedge B$ events with $Pr(A|B) > \frac{1}{2}$ and $Pr(B) \neq 0$. Therefore it is the case that $Pr(B|A) < \frac{1}{2}$.

Claim 6.3.4. Let $U = \mathbb{R}$ and $S \subseteq U$ be a well defined sample space with $A \wedge B$ events whilst both $Pr(B) \neq 0$ and $Pr(A) \neq 0$. Therefore it is the case that $Pr(B|A) + Pr(A|B) = 1$.

Claim 6.3.5. Let $U = \mathbb{R}$ and $S \subseteq U$ be a well defined sample space with $A \wedge B$ events whilst both $Pr(B) \neq 0$ and $Pr(A) \neq 0$. Therefore it is the case that $Pr(B|A) + Pr(B|A^c) = 1$.

Claim 6.3.6. Let $U = \mathbb{R}$ and $S \subseteq U$ be a well defined sample space with $A \wedge B$ events whilst both $Pr(B) \neq 0$ and $Pr(A) \neq 0$. Therefore it is the case that $Pr(B|A) + Pr(B^c|A) = 1$.

Claim 6.3.7. Let $U = \mathbb{R}$ and $S \subseteq U$ be a well defined sample space with $A \wedge B$ events with $Pr(A|B) > Pr(A)$ whilst both $Pr(B) \neq 0$ and $Pr(A) \neq 0$. Therefore it is the case that $Pr(B|A) < Pr(B)$.

Claim 6.3.8. Let $U = \mathbb{R}$ and $S \subseteq U$ be a well defined sample space with $A \wedge B$ events with $Pr(A|B) \geqslant Pr(A)$ whilst both $Pr(B) \neq 0$ and $Pr(A) \neq 0$. Therefore it is the case that $Pr(B|A) < Pr(B)$.

Claim 6.3.9. Let $U = \mathbb{R}$ and $S \subseteq U$ be a well defined sample space with $A \wedge B$ events with $Pr(A|B) = Pr(A)$ whilst both $Pr(B) \neq 0$ and $Pr(A) \neq 0$. Therefore it is the case that $Pr(B|A) = Pr(B)$.

Theorem 6.3.1. *Let S be a well defined sample space with $A \wedge B$ events where B is a non-trivial event. Therefore it is the case that $Pr(A) = Pr(A|B) \cdot Pr(B) + Pr(A|B^c) \cdot Pr(B^c)$*

For computational exercises: Perhaps use a Venn diagramme or Tree diagramme (where apropos); use the definition of the probability of W given M when $Pr(M) \neq 0$.

Exercise 6.3.1. This exercise is $\mathbb{B}$ of $\mathbb{D}$ 1.
Let S be a well-defined sample space with $A \wedge B$ events and we know that
$Pr(A) = 0.75, Pr(B) = 0.48$; and, $Pr(A \cup B) = 0.89$.

A. Find $Pr(A\|B)$	B. Find $Pr(B\|A)$	C. Find $Pr(A \cap B)$
D. Find $Pr(B^C \cup A)$	E. Find $Pr(A^C\|B)$	F. Find $Pr(B^C\|A)$
G. Find $Pr(A^C \cap B^C)$	H. Find $Pr(B^C \cap B)$	

Exercise 6.3.2. This exercise is $\mathbb{B}$ of $\mathbb{D}$ 1.
Let S be a well-defined sample space with $A_2 \wedge B_2$ events and assume
$Pr(A_2) = 0.4, Pr(B_2) = 0.3$; and, $Pr(A_2 \cap B_2) = 0.12$.
A. Find $Pr(B_2|A_2)$ B. Find $Pr(B_2^C|A_2^C)$ C. Find $Pr(A_2 \cap B_2)$
D. Find $Pr(B_2^C)$ E. Find $Pr(A_2^C|B_2)$ F. Find $Pr(B_2^C|A_2^C)$
G. Find $Pr(A_2^C \cap B_2^C)$

Exercise 6.3.3. This exercise is $\mathbb{B}$ of $\mathbb{D}$ 1.
Let S be a well-defined sample space with $A_3 \wedge B_3$ events. Assume
$Pr(A_3) = 0.4, Pr(B_3) = 0.2$; and, $Pr(A_3|B_3) = 0.2$
A. Find $Pr(A_3 \cap B_3)$ B. Find $Pr(B_3|A_3)$ C. Find $Pr(B_3^C)$
D. Find $Pr(A_3 \cup B_3)$ E. Find $Pr(B_3^C \cup B_3)$ E. Find $Pr(B_3|B_3)$

Exercise 6.3.4. This exercise is $\mathbb{B}$ of $\mathbb{D}$ 1.
Suppose a pair of dice is tossed. You view the sum of the two faces up-turned.

A. Find the probability that the sum of the sides facing up is more than 5 given the sum of the sides facing up is 7 (from here on out we will term this 'Find the probability that the you rolled more than 5 given that you rolled a 7.'

B. Find the probability that you rolled a seven given you rolled more than 5.

C. Find the probability that you rolled a less than 10 given you rolled more than 6.

Exercise 6.3.5. This exercise is $\mathbb{B}$ of $\mathbb{D}$ 1.
There is an urn. It contains 8 white, 3 red, 4 green, and 6 blue balls.

A. Two balls are drawn from the urn. Find the probability that both balls drawn are red.

B. Put all the balls back, shake up the urn and draw two balls (in succession) from the urn. Find the probability that the first ball is red and the second ball is green.

C. Put all the balls back, shake up the urn and draw two balls (in succession) from the urn. Find the probability that the second ball is green given the first ball is green.

D. Put all the balls back, shake up the urn and draw two balls (in succession) from the urn. Find the probability that the second ball is green given the first ball is not red.

E. Put all the balls back, shake up the urn and draw two balls (in succession) from the urn. Find the probability that one of the balls drawn is green and another is white.

Exercise 6.3.6. This exercise is $\mathbb{B}$ of $\mathbb{D}$ 1 but must be presented in conjunction with the previous exercise.
There is an urn. It contains 8 white, 3 red, 4 green, and 6 blue balls.

A. A ball is drawn from the urn. Note its colour; put it back and shake up the urn. Draw a ball from the urn. Find the probability that both balls drawn are red.

B. Put all the balls back, shake up the urn. A ball is drawn from the urn. Note its colour; put it back and shake up the urn. Draw a ball from the urn. Find the probability that the first ball is red and the second ball is green.

C. Put all the balls back, shake up the urn. A ball is drawn from the urn. Note its colour; put it back and shake up the urn. Draw a ball from the urn. Find the probability that the second ball is green given the first ball is green.

D. Put all the balls back, shake up the urn A ball is drawn from the urn. Note its colour; put it back and shake up the urn. Draw a ball from the urn. Find the probability that the second ball is green given the first ball is not red.

Exercise 6.3.7. This exercise is $\mathbb{B}$ of $\mathbb{D}$ 1.
There is an urn. It contains 8 white, 3 red, 4 green, and 6 blue balls.

A. Three balls are drawn from the urn. Find the probability that all of the balls drawn are red.

B. Put all the balls back, shake up the urn and draw three balls from the urn. Find the probability that one of the balls is red and the other two are not white.

C. Put all the balls back, shake up the urn and draw three balls from the urn. Find the probability that the at least two of the balls are blue.

D. Put all the balls back, shake up the urn and draw three balls from the urn. Find the probability that the at least one of the balls is green.

E. Put all the balls back, shake up the urn and draw three balls from the urn. Find the probability that the at least one of the balls is yellow.

6.4 Independence

Definition 6.4.1. Let U be a well defined universe and S be a well defined sample space (a subset of U). Let E and F be two events. E and F are **independent** if and only if it is the case that $Pr\,(F|E) = Pr\,(F)$ and $Pr\,(E|F) = Pr\,(E)$.

Alternately, the definition can be stated as:

Definition 6.4.2. Let U be a well defined universe and S be a well defined sample space (a subset of U). Let E and F be events. E and F are **independent** if and only if it is the case that $Pr\,(F \cap E) = Pr\,(F) \cdot Pr\,(E)$

Definition 6.4.3. Let U be a well defined universe and S be a well defined sample space (a subset of U). Let E, F, and G be three events. E, F, and G are **independent** if and only if it is the case that
(1) $Pr\,(G \cap F \cap E) = Pr\,(G) \cdot Pr\,(F) \cdot Pr\,(E)$
and
(2A) $Pr\,(F \cap E) = Pr\,(F) \cdot Pr\,(E)$
(2B) $Pr\,(G \cap E) = Pr\,(G) \cdot Pr\,(E)$
(2C) $Pr\,(F \cap G) = Pr\,(F) \cdot Pr\,(G)$

Definition 6.4.4. Let U be a well defined universe and S be a well defined sample space (a subset of U). Let E, F, G, and H be four events. E, F, G, and H are **independent** if and only if it is the case that
(1) $Pr\,(H \cap G \cap F \cap E) = Pr\,(H) \cdot Pr\,(G) \cdot Pr\,(F) \cdot Pr\,(E)$
and
(2) all triple-wise collections of events is satisfied such as (1) in the previous definition
(3) all pair-wise collections of events is satisfied such as (2) in the previous definition.

Definition 6.4.5. Let U be a well defined universe and S be a well defined sample space (a subset of U). Let Γ be a collection of events. Let Ω be the index set. The events are are **independent** if and only if it is the case that:
$Pr(\bigcap \Gamma) = Pr(\prod_{i \in \Omega} E_i)$ where $E_i \in \Omega\ \forall i \in \Omega \ldots$ throughout down to all triple-wise collections of events is satisfied such as in the previous definition and finally all pair-wise collections of events is satisfied as in the previous definitions.

The alternate definition of two events being independent, "let U be a well defined universe and S be a well defined sample space (a subset of U). Let E and F be events. E and F are independent if and only if it is the case that $Pr(F \cap E) = Pr(F) \cdot Pr(E)$" is the most useful one for theoretical work (proofs, counterexamples, etc. I opine) but the original is of use, too. Independence is a more generalised idea of 'reset.' Indeed consider:

Example 6.4.1. *There is an urn. It contains 8 white, 3 red, 4 green, and 6 blue balls.*

Non-independence exercise: *Shake up the urn and draw two balls (in succession) from the urn. Find the probability that the second ball is red is given the first ball is green.*

Independence exercise: *Put all the balls back, shake up the urn and draw a ball from the urn. Note its colour. Put the ball back and shake up the urn. Draw a ball from the urn. Find the probability that the second ball is red given the first ball is green.*

The idea is that when we have a well defined sample space and events A and B that are independent A does not effect B and B does not effect A (with regard to probability of one occurring given the other occurs or had occurred).

Example 6.4.2. *There is an urn. It contains 8 white, 3 red, 4 green, and 6 blue balls. Also, you have a fair six sided standard die. One chooses a ball from the urn and then rolls the die. One wishes to find the probability that he rolls an odd natural number on the die given he chose a blue ball. It is* $\frac{3}{6}$*. Note the probability of rolling an odd natural number on the die given he chose a blue ball is the same as the probability he rolls an odd natural number on the die.*
One rolls the die and then chooses a ball from the urn. One wishes to find the probability that he chooses a blue ball given he rolled an odd natural number on the die. It is $\frac{6}{21}$*. Note the the probability that he chooses a blue ball given he rolled an odd natural number on the die is the same as the probability he chooses a blue ball from the urn.*
So, the events are independent (and understood practically to be such).

Independence is a very critical property for applied probability and statistics; thus, I opine it is quite important to understand the concept for future course-work in this area of mathematics and for applied mathematics (physics, chemistry, computer science, actuarial science, technometrics, edumetrics, econometrics, biometrics, data analysis, finacial mathematics, etc.).

Definition 6.4.6. Let U be a well defined universe and S be a well defined sample space (a subset of U). Let E, F, and G be three events. E, F, and G are **locally independent** or **conditionally independent** if and only if it is the case that
(1) $Pr(F \cap E) = Pr(F) \cdot Pr(E)$
(2) $Pr(G \cap E) = Pr(G) \cdot Pr(E)$
(3) $Pr(F \cap G) = Pr(F) \cdot Pr(G)$

This concept generalises and is weaker (obviously) than independence. In applied areas there are other weaker conditions than even local independence (which is beyond the scope of the course).

6.5 Claims About Independence

Prove or disprove the claims (all are board-worthy):

Claim 6.5.1. Let $U = \mathbb{R}$ and $S \subseteq U$ be a well defined sample space with $E \wedge F$ independent events. Therefore it is the case that $E^c \wedge F$ are independent events.

Claim 6.5.2. Let $U = \mathbb{R}$ and $S \subseteq U$ be a well defined sample space with $E \wedge F$ independent events. Therefore it is the case that $E^c \wedge F^c$ are independent events.

Claim 6.5.3. Let $U = \mathbb{R}$ and $S \subseteq U$ be a well defined sample space with $E \wedge F$ independent events. Therefore it is the case that $E - F$ and $F - E$ are independent events.

Claim 6.5.4. Let $U = \mathbb{R}$ and $S \subseteq U$ be a well defined sample space with E an event. Therefore it is the case that E and E^c are independent events.

Claim 6.5.5. Let $U = \mathbb{R}$ and $S \subseteq U$ be a well defined sample space with E a non-trivial event. Therefore it is the case that E and E^c are independent events.

6.6 Bayes' Theorem

Shockingly, it is the case that one can find a primary (*a priori*) event probability given a secondary event has occurred under rigourous conditions. Such conditions are the premises of Bayes' Theorem; and if such are satisfied; then one can find a primary event probability given a secondary event has occurred.

Theorem 6.6.1. *(Bayes' Theorem). Let U be a well defined universe and S be a well defined sample space (a subset of U). Let the finite collection of events $\Omega = \{F_i | i \in \mathbb{N}\}$ be pair-wise mutually exclusive ($F_j \cap F_k = \emptyset \ni j \neq k$), non-trivial, and exhaustive ($\bigcup \Omega = S$ or even if one desires $\bigcup_{i=1}^{n} F_i = S$). Let E be a non-trivial event. Let $w \in \mathbb{N}_n$. It is the case that*

$$Pr\left(F_w | E\right) = \frac{Pr\left(F_w\right) \cdot Pr\left(E | F_w\right)}{\sum_{i=1}^{n} \left(Pr\left(F_i\right) \cdot Pr\left(E | F_i\right)\right)}$$

Corollary 6.6.1. *(Bayes' Theorem with a partition of size 2) Let U be a well defined universe and S be a well defined sample space (a subset of U). Let the collection of two events $\{F_1, F_2\}$ be mutually exclusive, non-trivial, and exhaustive ($F_1 \cup F_2 = S$). Let E be a non-trivial event. It is the case that*

$$Pr\left(F_1 | E\right) = \frac{Pr\left(F_1\right) \cdot Pr\left(E | F_1\right)}{\left(Pr\left(F_1\right) \cdot Pr\left(E | F_1\right) + \left(Pr\left(F_2\right) \cdot Pr\left(E | F_2\right)\right)\right.}$$

Corollary 6.6.2. *(Bayes' Theorem with a partition of size 3) Let U be a well defined universe and S be a well defined sample space (a subset of U). Let the collection of three events $\{F_1, F_2, F_3\}$ be pair-wise mutually exclusive, non-trivial, and exhaustive ($F_1 \cup F_2 \cup F_3 = S$). Let E be a non-trivial event. Let $a \in \mathbb{N}_3$. It is the case that*

$$Pr\left(F_a | E\right) = \frac{Pr\left(F_a\right) \cdot Pr\left(E | F_a\right)}{\left(Pr\left(F_1\right) \cdot Pr\left(E | F_1\right) + \left(Pr\left(F_2\right) \cdot Pr\left(E | F_2\right) + \left(Pr\left(F_3\right) \cdot Pr\left(E | F_3\right)\right)\right.\right.}$$

This concept generalises and is creates a branch of applied statistics called Bayesian statistics. I am not a fan of said. My major professor said the me that users of Bayesian statistics are more like magicians and snake oil salesmen than scientists (and he went on about smoke and mirrors, etc.) – so, my attitude is a creation of my ideas and from what I learnt from a person of whom I greatly respect.

6.7 Exercises About Bayes' Theorem

Exercise 6.7.1. Suppose we have an urn with 4 red, 1 blue, and 7 green balls. We draw out a ball look at it's colour. We draw another ball and look at it's colour.
A. What is the probability that the first ball is green?
B. What is the probability that the second ball is green given the first ball is blue?
C. What is the probability that the second ball is blue given the first ball is red?
D. What is the probability that the second ball is red given the first ball is red?
E. What is the probability that the second ball is blue given the first ball is blue?

Exercise 6.7.2. Suppose we have an urn with 4 red, 1 blue, and 7 green balls. We draw out a ball look at it's colour. We draw another ball and look at it's colour.
A. What is the probability that the second ball is red given the first ball is green?
B. What is the probability that the first ball is blue given the second ball is red?

Exercise 6.7.3. Suppose we have an urn with 4 red, 1 blue, and 7 green balls. We draw out a ball look at it's colour. We put it back; shake up the urn; and then we draw another ball and look at it's colour.
A. What is the probability that the second ball is red given the first ball is green?
B. What is the probability that the first ball is blue given the second ball is red?

Exercise 6.7.4. ($\mathbb{B}$) Prove Bayes' Theorem with a partition of size 2