23

Probability Trees and the Law of Multiplication

Exclusive and Exhaustive Events

Probability trees are one of the most useful tools in probability and applied mathematics. We will use them throughout the rest of this text. The graph in **Figure 1** is a probability tree. It has 4 vertices. The vertex X is the *root*. Every probability tree has exactly one root. The vertices A, B, and C are the *leaves*. In a probability tree there are arcs leaving each node that is not a leaf. These arcs have non-negative numerical labels such that the arcs leaving any node add up to 1 (the arcs will often have other labels as well). For example, in **Figure 1** the arc from node X to Node A the label .2 implies that on leaving vertex X the probability is .2 of going to Node A. The reason that the (numeric) labels on arcs leaving a node add to 1 is that the events being led up to are *exclusive* and *exhaustive*. They are exclusive is another way of saying they are disjoint. In this case that means that, P(AB) = P(AC) = P(BC) = 0. (Hence, P(A + B) = P(A) + P(B) = .5.) Since they add up to 1, that means that every possibility is accounted for: the events are exhaustive. After X; A, B, or C must occur but not any two of these. In general, a collection of events is exclusive and exhaustive if they are disjoint and their probabilities add to one.

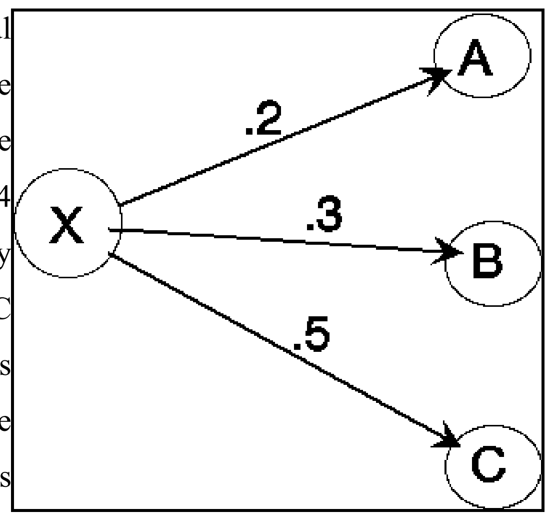

Figure 1 A Simple Probability Tree

The Law of Total Probability

Suppose that the events A, B, C are exclusive and exhaustive (as in **Figure 1**) then given any other event, Y, the law of total probability says that P(Y) = P(AY) + P(BY) + P(CY). More generally:

> **If the events A_1, A_2, through A_n are exclusive and exhaustive and X is any event, then: $P(X) = P(XA_1) + P(XA_2) + ... + P(XA_n)$.**

Law of Total Probability (First Version)

This is just one version of the law of total probability; later we will encounter it again. To paraphrase: given an arbitrary event, X, we can split it amongst any collection of exclusive and exhaustive events. The most important case is when we have an event, X, and its complement $\bar{X}$ (not X). X and $\bar{X}$ are by definition exclusive and exhaustive. In particular, given another event Y, by the law of total probability, we have $P(Y) = P(YX) + P(Y\bar{X})$. The law says: The probability that I work tomorrow is equal to the probability that it rains and I go to work, and the probability that it doesn't rain and I go to work.

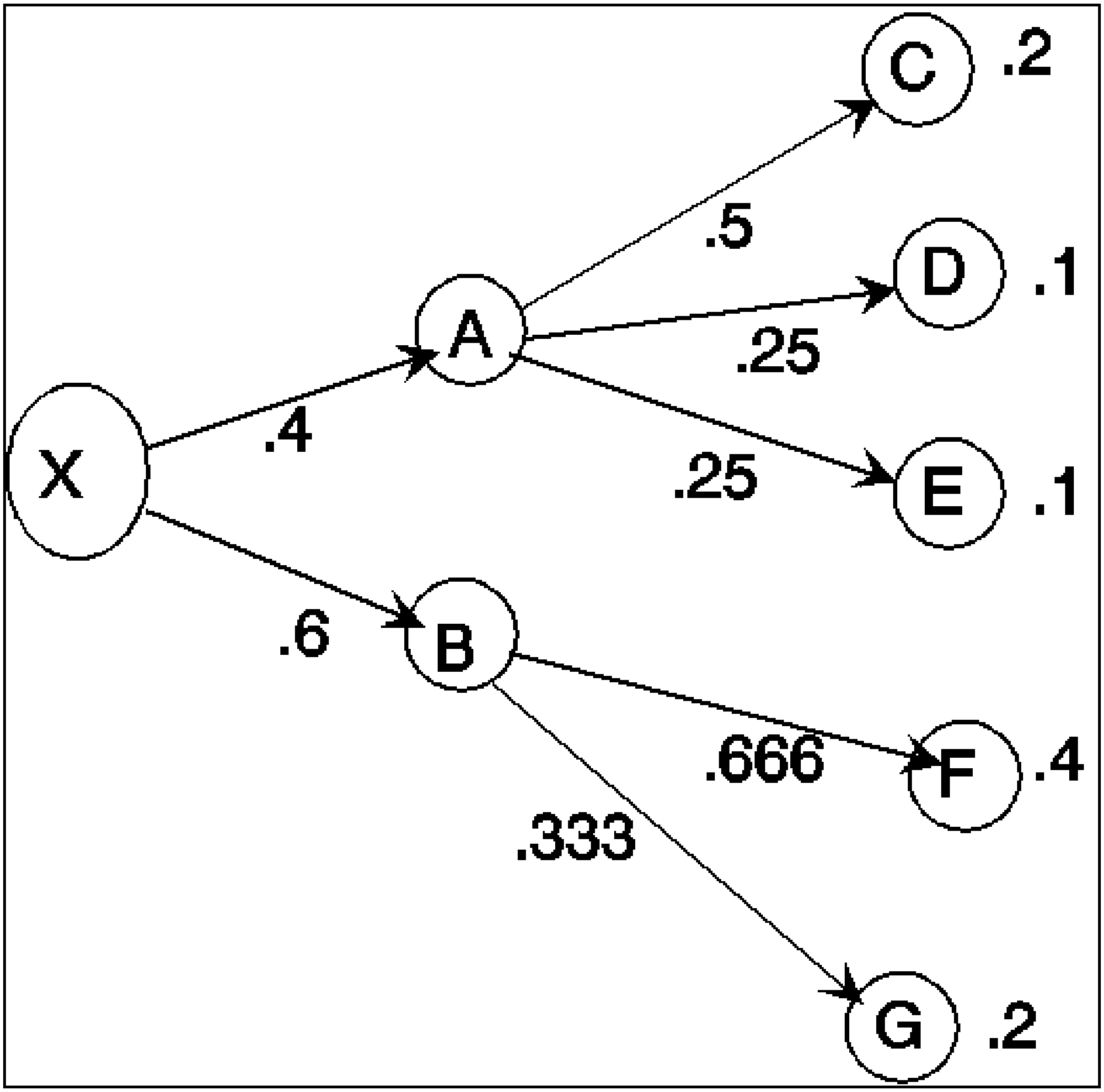

Figure 2 A Probability Tree With Five Leaves

The tree in **Figure 2** is also a probability tree. Notice that besides labels for the arcs, there are labels for each leaf. The label for a leaf is derived by multiplying the labels on each arc leading to that leaf. For example, the label on leaf D is the product of .4 and .25. Notice, that the labels on all the leaves add up to 1. This is always true. If it is not true for a given probability tree (that the labels on the nodes add up to 1) then a mistake has been made. A warning: if you later try to apply formulas to **Figure 2** you might need to write C as AC. In this case C = AC. C only occurs if A has occurred. C is the same as A and C. (Likewise, D = AD, E = AE, F = BF, G = BG.)

Conditional Probabilities

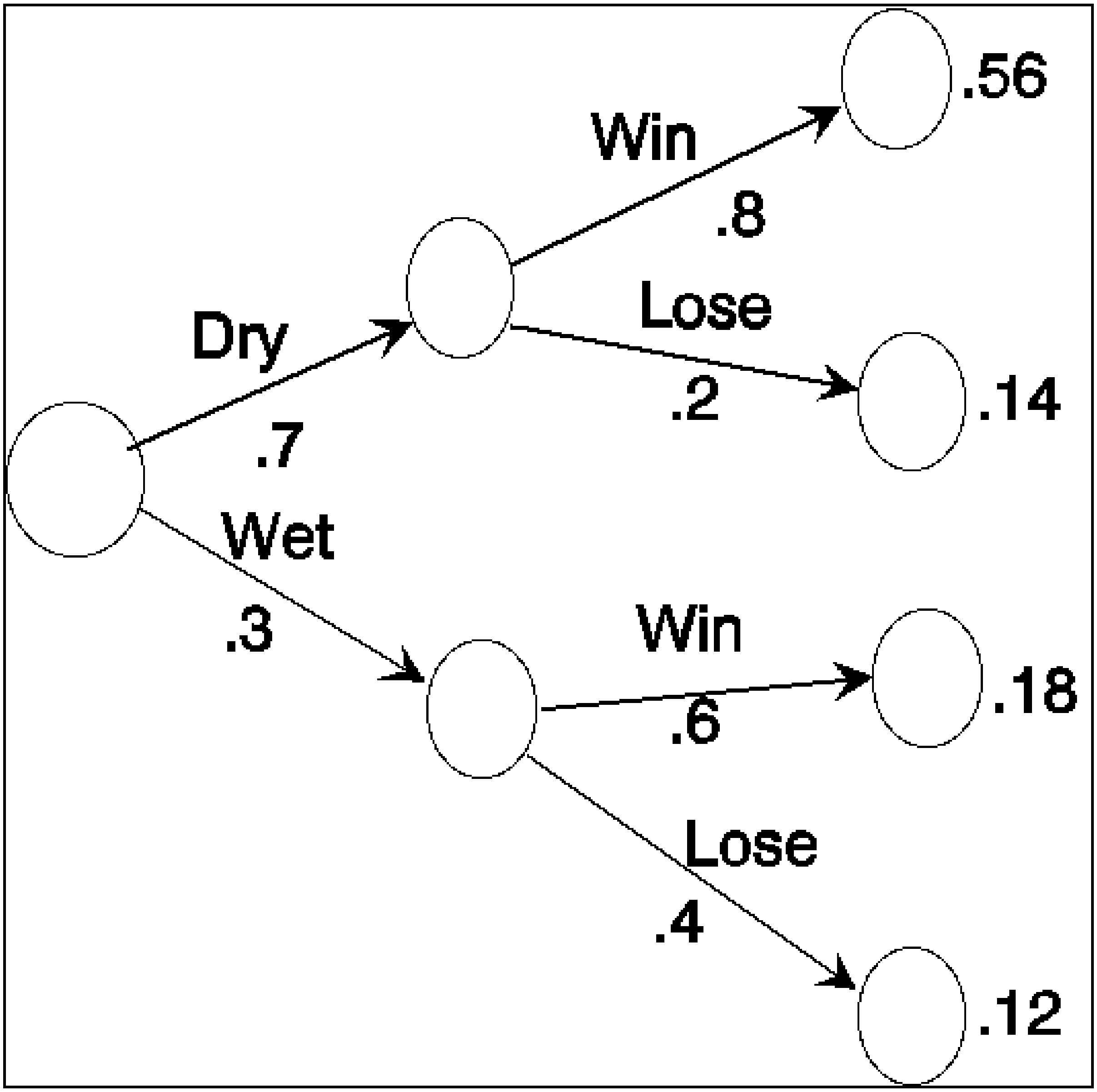

Figure 3 The Raiders: Commitment to Excellence

Let's consider a more specific example. The Raiders are a fair weather team. They play less well under wet conditions. This is illustrated in **Figure 3**.[1] The probability that the conditions are dry is .7. If so, the probabilities that the Raiders win is .8 and that they lose is .2. The probability they win when it is wet is .6 and that they lose is .4. By the law of total probability, the probability the Raiders win is .56 + .18 = .74. That is, by the law of total probability, the probability the Raiders win (.56) is the probability that it is dry and the Raiders win (.56) plus the probability that it is wet and the Raiders win (.18). You do not even have to think about the law of total probability. Just add the leaves corresponding to the Raiders winning. Notice how the probability tree gives you an easy way of processing complex information.

Conditional Probabilities

Strictly speaking the probabilities in **Figure 3** that follow the events wet and dry are *conditional* probabilities. We had to derive the unconditional probability P(win) = .74. The graph gives P(win|dry) = .8: the probability the raiders win **given** it is dry is .8. Read the vertical bar, |, as *given*. The graph also gives that the probability the raiders win given it is wet is .6: P(win| wet) = .6. Before giving the formal definition of conditional probabilities, let us reconsider the dice example:

1, 1	2, 1	3, 1	4, 1	5, 1	6, 1
1, 2	2, 2	3, 2	4, 2	5, 2	6, 2
1, 3	2, 3	3, 3	4, 3	5, 3	6, 3
1, 4	2, 4	3, 4	4, 4	5, 4	6, 4
1, 5	2, 5	3, 5	4, 5	5, 5	6, 5
1, 6	2, 6	3, 6	4, 6	5, 6	6, 6

Table 1 Sum of the Die is 5

[1]This is just a textbook example: the probability that the Raiders win is greater than this.

The event, S_5, that the sum of the die is 5, is 4 cases out of 36 so that $P(S_5) = 4/36 = 1/9$. That is an *unconditional* probability. We could ask a conditional question: Given that the first die is a 2, what is the probability that the sum is a 5? In other words, what is: $P(S_5|F_2)$? Now we are given that the first die is a 2; our universe is reduced to only one column of the original table.

2, 1
2, 2
2, 3
2, 4
2, 5
2, 6

Table 2 The First Die is a 2

Given that the first die is a 2, the case that the sum is a 5, is now 1 case out of 6 rather than 4 cases out of 36. Hence, $P(S_5|F_2) = 1/6$.

A key concept involved with conditional probabilities is **information**. In a way the heart of conditional probabilities and the upcoming concept of independence is information. In this case, the information that the first die is a 2 raised our estimate of the probability that the sum is a 5 from 1/9 to 1/6. We will later define two events as independent, if the knowledge of one event effects the probability of the other event occurring.

We define conditional probability of A given B as follows:

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

Dividing by the probability of B is where we reduce the size of our universe to reflect the given condition. In the above example, $P(S_5|F_2)$ is equal to $P(S_5 \cdot F_2)/P(F_2)$. Now $P(S_5 \cdot F_2)$ is the probability that the sum is 5 and the first die is 2. This is just one case out of 36 for a probability of 1/36. Altogether: $P(S_5|F_2) = P(S_5 \cdot F_2)/P(F_2) = (1/36)/(1/6) = 1/6$.

☐ **Exercise 1** What is the probability that the sum of two die is 4 given the first die is 2? That is solve: $P(S_4|F_2)$.

☐ **Exercise 2** Solve $P(S_4|F_4)$.

☐ **Exercise 3** Solve $P(S_7|F_2)$.

☐ **Exercise 4** Solve $P(F_2|S_4)$.

☐ **Exercise 5** Solve $P(F_4|S_4)$.

☐ **Exercise 6** Solve $P(F_6|S_2)$.

The Law of Multiplication

If we take the definition of P(A|B) and multiply both sides by P(B), we get the identity: P(AB) = P(A|B)·P(B). Similarly, we can show: P(AB) = P(B|A)·P(A). This is in fact, the law of multiplication:

For any two events A and B, P(A·B) = P(A|B)·P(B).

The Law of Multiplication

The Law of Total Probability Revisited

We can use the law of multiplication to rewrite the law of total probability.

If the events A_1, A_2, through A_n are exclusive and exhaustive and X is any event, then: **$P(X) = P(X|A_1)\cdot P(A_1) + P(X|A_2)\cdot P(A_2) + \ldots + P(X|A_n)\cdot P(A_n)$.**

The Law of Total Probability (Second Version)

When we used the graph in **Figure 3** to compute the unconditional probability that the Raiders win, we, in effect, used this law: P(Raiders win) = P(Raiders Win|it is dry)·P(It is dry) + P(Raiders win|it is wet)·P(it is wet).

Independence

Now we have come to the keystone of elementary probability. Mastery of this topic is the rite of passage into more serious probability. This topic should one day seem simple to you, and you will wonder why it took you so long to master it. You may blame the teacher and very likely the text (although surely the text can't be to blame). In fact, we have reached a place that requires a certain amount of digestion. You should linger here. Do not let impatience push you ahead too soon. By staying here until you have fully grasped the concept of independence, you will save yourself much time later. Mastery of this concept will better prepare you for applications later, such as **statistics**.

We will define independence three times! These definitions are equivalent.[1] That means if one definition is true all three are true. If one is false, they are all false. Anyone is sufficient to test for independence. This raises the question, why give all three definitions? In practice, one will be more useful than the others depending on the problem in question (and any rule could be the useful one). Using two rules is a good way to check your work. If you get different answers, you have made a mistake.

Events A and B are *independent* if one of the following is true:

1 $P(A|B) = P(A)$

2 $P(B|A) = P(B)$

3 $P(A{\cdot}B) = P(A){\cdot}P(B)$

Events A and B are *dependent* if they are not independent.

[1]The cases where they are not equivalent are when $P(A) = 0$ or $P(B) = 0$. These cases are trivial and uninteresting and can be ignored. I only mention them to keep the pedants happy.

The most common mistake I see when students test for independence is that the student assumes P(A·B) = P(A)·P(B). That is the student uses rule 1 or 2 to test for independence of two events, but the student assumes that rule 3 is satisfied. In effect, the student tests for independence of two events after assuming independence, and therefore automatically finds independence.

The assumption P(A·B) = P(A)·P(B) is often made because it is convenient for calculation and without any other justification. It is a most common mistake, and it has shown up in courtroom verdicts (later overturned) and all sorts of pseudo-scientific literature. It occurs in industry all the time.

Remember:

P(A·B) = P(A)·P(B)

if and only if A and B are independent!

☐ **Exercise 7** Prove that the three tests for independence given above are equivalent. That is, show that if one is true, then the others are true. If one is false, the others are false. [This exercise is more theoretical than the others and you might save it until you have finished the chapter and the other exercises.]

Example

S_6 and F_1 are dependent

1, 1	2, 1	3, 1	4, 1	5, 1	6, 1
1, 2	2, 2	3, 2	4, 2	5, 2	6, 2
1, 3	2, 3	3, 3	4, 3	5, 3	6, 3
1, 4	2, 4	3, 4	4, 4	5, 4	6, 4
1, 5	2, 5	3, 5	4, 5	5, 5	6, 5
1, 6	2, 6	3, 6	4, 6	5, 6	6, 6

Looking again at the fair throw of two fair die, let's consider the events *the sum is 6*, S_6, and *the first die is 1*, F_1. We can see from the table that $P(S_6) = 5/36$. However, we can also see $P(S_6|F_1) = 1/6$. That is, $P(S_6) \neq P(S_6|F_1)$ and hence S_6 and F_1 are dependent. However, it is easy to show that S_7 and F_1 are independent!

☐ **Exercise 8** Prove that S_7 and F_1 are independent.

Example

You call someone about puppies they advertised for sale in the newspaper. You are interested in getting a female puppy.[1] They have two puppies left. You ask *is one of the puppies a female?* They answer *yes*. Our question is, what is the probability that the other is a female. Logically, most people think that the answer should be 50%; and it would be, if the question asked were *is the puppy*

[1]This example is an old chestnut. The puppy version has appeared in Marilyn Vos Savant's column in *Parade* magazine that comes as a supplement in many Sunday newspapers. Ms. Savant periodically presents interesting problems. One probability problem created a large stir in the mathematics community because so many self-proclaimed "experts," all brandishing their Ph.D.'s, wrote to tell her she was wrong when in fact she was correct. In fact these self-proclaimed probability experts demonstrated that they did not understand conditional probabilities. Ms. Savant has since then become like these experts and has written foolish things about Andrew Wiles' proof of Fermat's "last theorem."

closest to you a female? But the correct answer is ⅓. One way to see it is by looking at the possibilities, which are:

1: MM

2: MF

3: FM

4: FF

By answering yes, they have reduced the possibilities to cases 2, 3 and 4. In only one of these is the other puppy a female. Another way of solving the problem is by application of the definition of conditional probability. We are asking *what is the probability that both puppies are female given that one is a female?* Let us denote these two events by FF and F1 respectively. We want P(FF| F1) and by definition that is $\frac{P(FF \cdot F1)}{P(F1)}$. The event FF·F1 is just the event FF (the only way that both puppies can be female **and** at least one be female is if they are both female.) But since we can assume that the genders of the two puppies are independent, P(FF) = ¼. By the law of addition the probability that at least one puppy is female is the probability that the first is female (½) plus the probability that the second is female (½) minus the probability that both are female (¼) to give a combined probability P(F1) = ¾. Plugging into the formula given for conditional probability, we get ⅓.

Independence and Information

Again independence is about information. Rule 1, P(A|B) = P(A), says that the occurrence of B says nothing about the probability of A occurring. Likewise rule 2, P(B|A) = P(B), says that the occurrence of A says nothing about B occurring. If two events are disjoint (exclusive) they do reveal information about each other, because if one occurs, that means the other can't occur. This is one of the most common mistakes that students make. They think that if events are disjoint that they are also independent.

☐ **Exercise 9** Use the definition of independence to prove that if A and B are disjoint (P(AB) = 0) then A and B are dependent. Ignore the trivial cases where either P(A) = 0 or P(B) = 0.

If two (non-empty) events are disjoint, they are dependent!

If we consider the Raiders example, **Figure 3**, the Raiders are more likely to win if it is dry. If you are a gambler, and you have the information that the weather will be dry, it can change your willingness to bet on the outcome of the game. The events that the weather is dry and the Raiders win are **dependent**.

	D	E	F
A	0	.2	.1
B	.1	.2	.2
C	.1	.1	0

The table above is a table of joint probabilities. It is an exclusive and exhaustive table in that the entries are all disjoint and add to 1. Each entry is the joint probability of its row and column. For example, P(BD) is the number in the B row and D column, .1. The event A is the first row, so its probability is .3. The event E is the second column, so its probability is .5.

☐ **Exercise 10** What is P(A|D)? What is P(A|E), P(A|F), P(A|B)?

☐ **Exercise 11** Are A and B independent? How about A and D? A and E? A and F?

☐ **Exercise 12** What is P(B|D)? What is P(B|E), P(B|F)?

☐ **Exercise 13** Are B and C independent? How about B and D? B and E? B and F?

☐ **Exercise 14** What is P(C|D)? What is P(C|E), P(C|F)?

☐ **Exercise 15** Are C and D independent? How about C and E? C and F?

For the following exercises, assume that we make one fair throw of a fair die. Let E be the event the face is even. Let F_i be the event that the face is i. Let L_i be the event that the face is less than or equal to i. Note that $P(L_i) = i/6$. For example, L_2 is the event that the face is 1 or 2, and therefore $P(L_2) = 2/6 = 1/3$.

☐ **Exercise 16** Are E and F_3 independent?

☐ **Exercise 17** Are E and F_4 independent?

☐ **Exercise 18** Are E and L_3 independent?

☐ **Exercise 19** Are E and L_2 independent?

☐ **Exercise 20** Are L_4 and L_2 independent?

☐ **Exercise 21** Are L_4 and L_6 independent?

On Fair Throws and an Unsolvable Problem

(This Section Optional)

Throughout the examples I make mention of fair throws of a fair coin (or fair die). It is generally assumed that the reader knows what this means. Now, the reader can better appreciate it because fair tosses are *independent* tosses.

It is frequently a first lesson in probability classes to raise the following question: *A man flips a coin ten times in a row and gets heads each time. What is the probability that he throws heads on the eleventh throw?* The stock answer is ½, followed by the cliché that *the coin has no memory*. The incorrect answer, that the professor is presumably waiting for, is for some student to say, 1/2048: the probability of 11 heads in a row.

The trouble with the preceding scenario is that ½ is the correct answer only if it is stipulated that the coin is fair and is fairly tossed. The student who answers that the probability should be 1/2048 is thinking in terms of a *law of averages*. The problem is there is no law of averages in probability; what is called the law of averages is a particularly bad misinterpretation of the law of large numbers.[1]

If we were to see a man toss a coin ten heads in a row and to get ten heads, wouldn't it be fair to suspect that the coin and/or the throw is unfair? In this more general (and more lifelike) problem, what is our estimate that he throws heads on the eleventh throw? This is a fundamental question of probability and statistics (since it calls for a decision based upon observed data, we are now crossing into statistics). Intuitively we should place the probability that he throws heads again as pretty high. One solution technique says that answer should be 11/12, and I would place that as a lower bound for the answer.[2] The fact is that the question is somewhat philosophical, and there probably will never be a universally agreed on answer. If the man who is tossing the coin bets money on the outcome of the eleventh throw, I would bet with him. If he bets $100 that the eleventh throw is tails, I would ride along on that bet if I could. This problem is a simplified version of life and death problems facing the military. I know; I have on occasion worked on such problems.

[1]The law of large numbers (LLN) might be better described as a law of averages, but for whatever the reason, it is not. The law of averages seems to be used almost exclusively to mean a fallacious misinterpretation of the LLN. The LLN is an important topic but is not included in this text because I can't include everything. Having read this text you should be in good shape to study it elsewhere.

[2]The technique I refer to is Laplace's law of succession. For a discussion see *An Introduction to Probability Theory and Its Applications*, Vol. 1, 3rd. ed. by William Feller; Wiley, 1968.

1. $$P(S_4|F_2) = \frac{P(S_4 \cdot F_2)}{P(F_2)} = \frac{\frac{1}{36}}{\frac{1}{6}} = \frac{1}{6}$$

2. $$P(S_4|F_4) = \frac{P(S_4 \cdot F_4)}{P(F_4)} = \frac{0}{\frac{1}{6}} = 0$$

3. $$P(S_7|F_2) = \frac{P(S_7 \cdot F_2)}{P(F_2)} =7 \frac{\frac{1}{36}}{\frac{1}{6}} = \frac{1}{6}$$

4. $$P(F_2|S_4) = \frac{P(F_2 \cdot S_4)}{P(S_4)} = \frac{\frac{1}{36}}{\frac{3}{36}} = \frac{1}{3}$$

5. $$P(F_4|S_4) = \frac{P(F_4 \cdot S_4)}{P(S_4)} = \frac{0}{\frac{3}{36}} = 0$$

6. $$P(F_6|S_2) = \frac{P(F_6 \cdot S_2)}{P(S_2)} = \frac{0}{\frac{1}{36}} = 0$$

7. We need to show that the following conditions are equivalent. First we will assume that neither P(A) = 0 or P(B) = 0 in which case P(B|A) and P(A|B) would not be defined respectively.
   1. P(A|B) = P(A)
   2. P(B|A) = P(B)
   3. P(A·B) = P(A)·P(B)

   We will prove equivalence by showing 1⇒3⇒2 and 3⇒1. (This scheme also implies that 2⇒1.) We start by assuming 1. P(A|B) = P(AB)/P(B) = P(A). Multiply both sides of the second equality by P(B) and we get 3. Divide both sides of 3 by P(A) and we get 2. Divide both sides of 3 by P(B) and we get 1.

8. This can be proven any **one** of three different ways!!

   1: $P(S_7 | F_1) = P(S_7) = 1/6$.

   2: $P(F_1 | S_7) = P(F_1) = 1/6$.

   3: $P(S_7 \cdot F_1) = P(S_7) \cdot P(F_1) = 1/36$.

9. $$P(A|B) = \frac{P(A \cdot B)}{P(B)} = \frac{0}{P(B)} = 0 \neq P(A)$$

10. 0; .4; ⅓; 0

11. Remember any one of these problems can be tested any one of three ways!
    no (see exercise 9); no (same reason); no; no.

12. .5; .4; ⅔

13. no; yes; no; no

14. .5; .2; 0

15. no; yes; no

16. No: $P(E) = ½$; $P(F_3) = 1/6$; but $P(E \cdot F_3) = 0$.

17. No: $P(E) = ½$; $P(F_4) = 1/6$; but $P(E \cdot F_4) = 1/6$.

18. No: $P(E) = ½$; $P(L_3) = ½$; but $P(E \cdot L_3) = 1/6$.

19. Yes: $P(E) = ½$; $P(L_2) = ⅓$; $P(E \cdot L_2) = 1/6$.

20. No: $P(L_4) = ⅔$; $P(L_2) = ⅓$; but $P(L_4 \cdot L_2) = ⅓$.

21. Yes: $P(L_4) = ⅔$; $P(L_6) = 1$; $P(L_4 \cdot L_6) = ⅔$.