Reminder of some probability rules:
- Complement rule: \( P(E^c) = 1 - P(E) \)
- General addition rule: \( P(E \cup F) = P(E) + P(F) - P(E \text{ and } F) \)
  - For disjoint events: \( P(E \cup F) = P(E) + P(F) \)
- Conditional probability: \( P(E | F) = \frac{P(E \cap F)}{P(F)} \)
- General multiplication rule: \( P(E \cap F) = P(E) \times P(F | E) \)
  - For independent events: \( P(E \cap F) = P(E) \times P(F) \)
  - For many independent events: \( P(E_1 \cap E_2 \cap \ldots \cap E_k) = P(E_1) \times P(E_2) \times \ldots \times P(E_k) \)

Now we’ll use what we already know to discover two new rules:
- **Law of total probability**, for finding an unconditional probability from conditional ones
- **Bayes’ rule**, for finding “reverse” conditional probabilities

The key to deriving and applying these rules will be probability tables and probability trees.

**Example 5-1: Document Errors**
Suppose that an office employs three associates who prepare a certain kind of document. Delia prepares 60% of these documents, Francis 30%, and Gino 10%. Delia makes an error in 15% of the documents that she prepares, Francis in 25%, and Gino in 5%.

a) Translate the given information into probability statements, using appropriate symbols.

b) Suppose that the given percentages hold exactly for a population of 1000 documents. Fill in the table below. [Hint: Start with the total number of documents prepared by each person. Then proceed to the error/not breakdown for each person.]

<table>
<thead>
<tr>
<th></th>
<th>Contains an error</th>
<th>Does not contain an error</th>
<th>Total documents</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prepared by Delia</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prepared by Francis</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prepared by Gino</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total documents</td>
<td></td>
<td></td>
<td>1000</td>
</tr>
</tbody>
</table>

d) Report the probability that a randomly selected document is both prepared by Delia and contains an error. Also use appropriate symbols to represent this probability.

e) Indicate how the probability in d) could have been found directly from the multiplication rule.
f) Use the table to determine the (unconditional) probability that a randomly selected document contains an error.

g) Indicate how the probability in f) could have been found directly from the given probabilities.

- **Law of total probability**: If $E_1, E_2, \ldots, E_k$ are disjoint events that comprise the entire sample space of possibilities, then $P(F) = \sum [P(F | E_i) \times P(E_i)]$.
  - Notice that this rule allows for calculating an unconditional probability by taking a *weighted average* of the conditional probabilities.

h) Now suppose that a randomly selected document is found to contain an error. Use the table to determine the updated (conditional) probability that Delia prepared it. Also use appropriate symbols to represent this probability.

i) Indicate how the probability in h) could have been found directly from the given probabilities.

- **Bayes’ rule**: If $E_1, E_2, \ldots, E_k$ are disjoint events that comprise the entire sample space of possibilities, then $P(E_m | F) = [P(F | E_m) \times P(E_m)] / \sum [P(F | E_i) \times P(E_i)]$.
  - Notice that this rule allows for calculating a “reverse” conditional probability.

j) Continue to suppose that a randomly selected document is found to contain an error. Determine the updated (conditional) probability that Francis prepared it. Then determine the updated (conditional) probability that Gino prepared it.

k) What do you notice about the three updated probabilities?

l) For each of the three associates, compare the (prior) probability that he/she prepared the document to the updated (conditional) probability given that the document contains an error.

\[
\text{Delia:} \\
\text{Francis:} \\
\text{Gino:}
\]
m) Given that a randomly selected document is found to contain an error, who is most likely to have prepared it? Who is least likely? Explain why these make sense.

n) Show how to use a probability tree to represent the given probabilities and to determine the other probabilities that you have calculated.

Example 5-2: AIDS Testing
The ELISA test for AIDS was used in the screening of blood donations in the 1990s. As with most medical diagnostic tests, the ELISA test is not infallible. If a person actually carries the AIDS virus, experts estimate that this test gives a (correct) positive result 97.7% of the time. If a person does not carry the AIDS virus, ELISA gives a (correct) negative result 92.6% of the time. Experts also estimate that 0.5% of the American public carries the AIDS virus.

a) Suppose that a randomly selected American tells you that they have tested positive. Given this information, how likely do you think it is that the person actually carries the AIDS virus? Make a prediction for this conditional probability without performing any calculations.

To determine this probability, imagine a hypothetical population of 1,000,000 people for whom these percentages hold exactly.

b) Assuming that 0.5% of the population of 1,000,000 people carries AIDS, how many such carriers are there in the population? How many non-carriers are there? (Record these in the table below.)

c) Consider for now just the carriers. If 97.7% of them test positive, how many test positive? How many carriers does that leave who test negative? (Record these in the table.)

d) Now consider only the non-carriers. If 92.6% of them test negative, how many test negative? How many non-carriers does that leave who test positive? (Record these in the table.)
e) Determine the total number of positive test results and the total number of negative test results. (Record these in the table.)

<table>
<thead>
<tr>
<th></th>
<th>Positive test</th>
<th>Negative test</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carries AIDS virus</td>
<td>(c)</td>
<td>(c)</td>
<td>(b)</td>
</tr>
<tr>
<td>Does not carry AIDS</td>
<td>(d)</td>
<td>(d)</td>
<td>(b)</td>
</tr>
<tr>
<td>Total</td>
<td>(e)</td>
<td>(e)</td>
<td>1,000,000</td>
</tr>
</tbody>
</table>

f) Of those who test positive, what proportion actually carry the disease? How does this compare to your prediction in a)? Explain why this probability is smaller than most people expect.

g) Of those who test negative, what proportion do not have the disease? Can you be very confident that those who are allowed to give blood (because they test negative) are not giving AIDS-infected blood?

Example 5-3: “Monty Hall” Problem
Suppose that on a game show a new car is hidden behind one door, while goats are hidden behind two other doors. A contestant picks a door, and then the host (to build suspense) reveals what’s behind a different door that he knows to have a goat. Then the host asks whether the contestant prefers to stick with the original door or switch to the remaining door.

a) Which strategy would you pick: stay or switch? Or do you think it doesn’t matter? Explain.

b) Let’s figure out the conditional probabilities by imagining 3,000 plays of the game, assuming that you initially select door #1. Fill in the following table (start with the rightmost column and then the “behind door #3 row):

<table>
<thead>
<tr>
<th></th>
<th>Host reveals door #2</th>
<th>Host reveals door #3</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Car is behind door #1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Car is behind door #2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Car is behind door #3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td>3,000</td>
</tr>
</tbody>
</table>

c) Given that the host reveals door #2, what is the conditional probability that the car is really behind door #3? Given that the host reveals door #3, what is the conditional probability that the car is really behind door #2? What does this mean in terms of the “stay or switch” decision?

d) Explain why the optimal strategy makes sense.
e) Now suppose that you analysis of the history of the game show suggests that the car is placed behind door 1 50% of the time, behind door 2 40% of the time, and behind door 3 only 10% of the time. What is the optimal strategy (which door should you pick, and should you stay or switch)? What is the probability of winning the car with this strategy? Explain your answer.

Example 5-4: Finding a Bug
Suppose that a computer program has a “bug” that is equally likely to be in any one of 3 sections of code. Call these sections A, B, and C, and assume that the program only has one bug. Now suppose that when you search a particular section to look for the bug, the probability is .9 that you will find the bug if it really is in that section (so the probability is .1 that you will miss the bug even when it is in the section that you search). Also assume that you never find a bug that is not really there.

Finally, suppose that you search section B and do not find the bug. Use Bayes’ rule to determine the updated probabilities, given this evidence, that the bug is in section A, B, C. [Hint: Start by clearly identifying the events of interest. You might use a probability tree.]