Reminder of some probability rules:

- Complement rule: \( P(\text{not } E) = 1 - P(E) \)
- General addition rule: \( P(E \text{ or } F) = P(E) + P(F) - P(E \text{ and } F) \)
  - For disjoint events: \( P(E \text{ or } F) = P(E) + P(F) \)
- Conditional probability: \( P(E \mid F) = \frac{P(E \text{ and } F)}{P(F)} \)
- General multiplication rule: \( P(E \text{ and } F) = P(E) \times P(F \mid E) \)
  - For independent events: \( P(E \text{ and } F) = P(E) \times P(F) \)
  - For many independent events: \( P(E_1 \text{ and } E_2 \text{ and } \ldots \ E_k) = P(E_1) \times P(E_2) \times \ldots \times P(E_k) \)

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 11-1: 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 10% 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>Prepared by Delia</th>
<th>Contains an error</th>
<th>Does not contain an error</th>
<th>Total documents</th>
</tr>
</thead>
<tbody>
<tr>
<td>( E_D )</td>
<td></td>
<td></td>
<td>600</td>
</tr>
<tr>
<td>( E_F )</td>
<td></td>
<td></td>
<td>300</td>
</tr>
<tr>
<td>( E_G )</td>
<td></td>
<td></td>
<td>100</td>
</tr>
<tr>
<td>Total documents</td>
<td></td>
<td></td>
<td>1000</td>
</tr>
</tbody>
</table>

C) Convert this table into a probability table by dividing all of the counts by 1000:

<table>
<thead>
<tr>
<th>Prepared by Delia</th>
<th>Contains an error</th>
<th>Does not contain an error</th>
<th>Total documents</th>
</tr>
</thead>
<tbody>
<tr>
<td>( E_D )</td>
<td></td>
<td></td>
<td>0.600</td>
</tr>
<tr>
<td>( E_F )</td>
<td></td>
<td></td>
<td>0.300</td>
</tr>
<tr>
<td>( E_G )</td>
<td></td>
<td></td>
<td>0.100</td>
</tr>
<tr>
<td>Total documents</td>
<td></td>
<td></td>
<td>1.000</td>
</tr>
</tbody>
</table>
d) Using either of these tables, 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 either 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 either 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.

   Delia:

   Francis:

   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 11-2: Seat Belts**
A recent Maryland highway safety study found that the driver was wearing a seat belt in 77% of
all accidents. Accident reports indicated that 92% of those drivers escaped serious injury
(hospitalization or death), compared to 63% of the drivers who were not wearing a seat belt.

a) Use a probability tree to represent the given probabilities.
b) Determine the probability that a randomly selected accident driver was not wearing a seat belt and was seriously injured. Also indicate the probability rule that applies here, and use appropriate symbols.

c) Determine the probability that a randomly selected accident driver was seriously injured, without regard for whether or not the driver was wearing a seat belt. Also indicate the probability rule that applies here, and use appropriate symbols. Finally, show how this calculation is a weighted average.

d) Determine the conditional probability that a randomly selected driver was not wearing a seat belt, given that he/she was seriously injured. Also determine the conditional probability that the driver was wearing a seat belt, given that he/she was seriously injured. Which possibility is more likely? Indicate the probability rule that applies here, and use appropriate symbols.

e) Now suppose that the randomly selected driver is not seriously injured. Determine the conditional probability that the driver was wearing a seat belt, and also determine the conditional probability that the driver was not wearing a seat belt. Comment on how these probabilities compare to the (unconditional) probabilities of wearing a seat belt or not.

f) Show how to use a probability table to answer these questions.
Example 11-3: 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 positive result 97.7% of the time. If a person does not carry the AIDS virus, ELISA gives a negative result 92.6% of the time. Experts also estimate that 0.5% of the American public carries the AIDS virus.

a) Suppose that someone 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 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 11-4: “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 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) We will investigate these strategies by playing the game several times with both strategies, using the site: http://www.shodor.org/interactivate/activities/monty3/. Record the results below:

<table>
<thead>
<tr>
<th></th>
<th>“Stay” strategy</th>
<th>“Switch” strategy</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wins</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Losses</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

c) In what proportion of games employing the “stay” strategy did the contestant win? What was this proportion for the “switch” strategy? Does one strategy appear to be superior to the other?

d) Now we’ll try to figure out the exact conditional probabilities by imagining 3,000 plays of the game, assuming that you initially select door #1. Fill in the following table:

<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>

e) 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?

f) Explain why the optimal strategy makes sense.