### 3.5 Problems

1. In a certain population of males, the following longevity probabilities are determined.
   
   - \( P(\text{Live to age 60}) = 0.90 \)
   - \( P(\text{Live to age 70, given live to age 60}) = 0.80 \)
   - \( P(\text{Live to age 80, given live to age 70}) = 0.75 \)
   
   From this information, calculate the following probabilities.
   
   - \( P(\text{Live to age 70}) \)
   - \( P(\text{Live to age 80}) \)
   - \( P(\text{Live to age 80, given live to age 60}) \)

2. Refer to the “barking dogs” problem in section 3.2.
   
   (a) Are the events “Angel barks” and “Brutus barks” **statistically independent**?
   
   (b) Calculate each of the following probabilities.
   
   - \( P(\text{Angel barks OR Brutus barks}) \)
   - \( P(\text{NEITHER Angel barks NOR Brutus barks}) \)
   - \( P(\text{Angel does not bark AND Brutus does not bark}) \)
   - \( P(\text{Only Angel barks}) \)
   - \( P(\text{Angel barks AND Brutus does not bark}) \)
   - \( P(\text{Only Brutus barks}) \)
   - \( P(\text{Angel does not bark AND Brutus barks}) \)
   - \( P(\text{Exactly one dog barks}) \)
   - \( P(\text{Brutus barks | Angel barks}) \)
   - \( P(\text{Brutus does not bark | Angel barks}) \)
   - \( P(\text{Angel barks | Brutus does not bark}) \)

   Also construct a Venn diagram, and a 2 × 2 probability table, including marginal sums.

3. Referring to the “urn model” in section 3.2, are the events \( A = \text{“First ball is red”} \) and \( B = \text{“Second ball is red”} \) independent in this sampling **without replacement** scenario? Does this agree with your intuition? Rework this problem in the sampling **with replacement** scenario.

4. After much teaching experience, Professor F has come up with a conjecture about office hours: “There is a 75% probability that a random student arrives to a scheduled office hour within the first 15 minutes (event \( A \)), from among those students who come at all (event \( B \)). Furthermore, there is an 80% probability that no students will come to the office hour, given that no students arrive within the first 15 mins.” Answer the following. (Note: Some algebra may be involved.)
   
   (a) Calculate \( P(B) \), the probability that any students come to the office hour.
   
   (b) Calculate \( P(A) \), the probability that any students arrive in the first 15 mins of the office hour.
   
   (c) Sketch a Venn diagram, and label all probabilities in it.
5. Suppose that, in a certain population of cancer patients having similar ages, lifestyles, etc., two categorical variables – $I =$ Income (Low, Middle, High) and $J =$ Disease stage (1, 2, 3, 4) – have probabilities corresponding to the column and row marginal sums in the $3 \times 4$ table shown.

<table>
<thead>
<tr>
<th>Income Level</th>
<th>Cancer stage 1</th>
<th>Cancer stage 2</th>
<th>Cancer stage 3</th>
<th>Cancer stage 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td></td>
<td></td>
<td></td>
<td>0.5</td>
</tr>
<tr>
<td>Middle</td>
<td></td>
<td></td>
<td></td>
<td>0.3</td>
</tr>
<tr>
<td>High</td>
<td>0.1</td>
<td>0.2</td>
<td>0.3</td>
<td>0.4</td>
</tr>
<tr>
<td></td>
<td>0.2</td>
<td>0.3</td>
<td>0.4</td>
<td>1.0</td>
</tr>
</tbody>
</table>

(a) Suppose $I$ and $J$ are **statistically independent**. Complete all entries in the table.

(b) For each row $i = 1, 2, 3$, calculate the following conditional probabilities, across the columns $j = 1, 2, 3, 4$:

- $P(\text{Low Inc } | \text{ Stage } j)$
- $P(\text{Mid Inc } | \text{ Stage } j)$
- $P(\text{High Inc } | \text{ Stage } j)$

Confirm that, for $j = 1, 2, 3, 4$:

- $P(\text{Low Income } | \text{ Stage } j)$ are all equal to the unconditional row probability $P(\text{Low Income})$.
- $P(\text{Mid Income } | \text{ Stage } j)$ are all equal to the unconditional row probability $P(\text{Mid Income})$.
- $P(\text{High Income } | \text{ Stage } j)$ are all equal to the unconditional row probability $P(\text{High Income})$.

That is, $P(\text{Income } i \mid \text{ Stage } j) = P(\text{Income } i)$. Is this consistent with the information in (a)? Why?

(c) Now for each column $j = 1, 2, 3, 4$, compute the following conditional probabilities, down the rows $i = 1, 2, 3$:

Likewise confirm that, for $i = 1, 2, 3$:

- $P(\text{Stage } j \mid \text{ Income } i)$ are all equal to the unconditional column probability $P(\text{Stage } j)$. That is, $P(\text{Stage } j \mid \text{ Income } i) = P(\text{Stage } j)$. Is this consistent with the information in (a)? Why?

* Technically, we have only defined statistical independence for events, but it can be formally extended to general random variables in a natural way. For categorical variables such as these, every category (viewed as an event) in $I$, is **statistically independent** of every category (viewed as an event) in $J$, and vice versa.
6. A certain medical syndrome is usually associated with two overlapping sets of symptoms, $A$ and $B$. Suppose it is known that:

- If $B$ occurs, then $A$ occurs with probability 0.80.
- If $A$ occurs, then $B$ occurs with probability 0.90.
- If $A$ does not occur, then $B$ does not occur with probability 0.85.

Find the probability that $A$ does not occur if $B$ does not occur. (Hint: Use a $2 \times 2$ probability table; label the marginal probabilities $a = P(A)$, $b = P(B)$, and the intersection probability $c = P(A \cap B)$. Fill out the rest of the table from the given statements using these symbols, then use some algebra.)

7. The progression of a certain disease is typically characterized by the onset of up to three distinct symptoms, with the following properties:

- Each symptom occurs with 60% probability.
- If a single symptom occurs, there is a 45% probability that the two other symptoms will also occur.
- If any two symptoms occur, there is a 75% probability that the remaining symptom will also occur.

Answer each of the following. (Hint: Use a Venn diagram.)

(a) What is the probability that all three symptoms will occur?
(b) What is the probability that at least two symptoms occur?
(c) What is the probability that exactly two symptoms occur?
(d) What is the probability that exactly one symptom occurs?
(e) What is the probability that none of the symptoms occurs?
(f) Is the event that a symptom occurs statistically independent of the event that any other symptom occurs?

8. I have a nephew Berkeley and niece Chelsea (true) who, when very young, would occasionally visit their Uncle Ismor on weekends (also true). Furthermore,

i. Berkeley and Chelsea visited independently of one another.

ii. Berkeley visited with probability 80%.

iii. Chelsea visited with probability 75%.

However, it often happened that some object in his house – especially if it was fragile – accidentally broke during such visits (not true). Furthermore,

iv. The probability of such an accident occurring, given that both children visited, was 90%.

v. The probability of such an accident occurring, given that only Berkeley visited, was 35%.

vi. The probability of such an accident occurring, given that only Chelsea visited, was 20%.

vii. The probability of such an accident occurring, given that neither child visited, was 2%.

Sketch and label a Venn diagram for events $A =$ Accident, $B =$ Berkeley visited, and $C =$ Chelsea visited. (Hint: The Exercise on page 3.2-3 might be useful.)
9. At a certain meteorological station, data are being collected about the behavior of thunderstorms, using two lightning rods $A$ and $B$. It is determined that, during a typical storm, there is a 99% probability that lightning will strike at least one of the rods. Moreover, if $A$ is struck, there is a 60% probability that $B$ will also be struck, whereas if $B$ is struck, there is a 75% probability that $A$ will also be struck. Calculate the probability of each of the following events. 

*(Hint: See PowerPoint section 3.2, slide 30.)*

- Both rods $A$ and $B$ are struck by lightning
- Rod $A$ is struck by lightning
- Rod $B$ is struck by lightning

Are the two events “$A$ is struck” and “$B$ is struck” statistically independent? Explain.

10. The Monty Hall Problem (simplest version)

Between 1963 and 1976, a popular game show called “Let’s Make A Deal” aired on network television, starring charismatic host Monty Hall, who would engage in “deals” – small games of chance – with randomly chosen studio audience members (usually dressed in outrageous costumes) for cash and prizes. One of these games consisted of first having a contestant pick one of three closed doors, behind one of which was a big prize (such as a car), and behind the other two were “zonk” prizes (often a goat, or some other farm animal). Once a selection was made, Hall – who knew what was behind each door – would open one of the other doors that contained a zonk. At this point, Hall would then offer the contestant a chance to switch their choice to the other closed door, or stay with their original choice, before finally revealing the contestant’s chosen prize.

*Question: In order to avoid “getting zonked,” should the optimal strategy for the contestant be to switch, stay, or does it not make a difference?*
11. (a) Given the following information about three events $A$, $B$, and $C$.

\[
\begin{align*}
P(A \cup B) &= 0.69 & P(A \cap B) &= 0.19 \\
P(A \cup C) &= 0.70 & P(A \cap C) &= 0.20 \\
P(B \cup C) &= 0.71 & P(B \cap C) &= 0.21 
\end{align*}
\]

Find the values of $P(A)$, $P(B)$, and $P(C)$.

(b) Suppose it is also known that the two events $A \cap C$ and $B$ are statistically independent. Sketch a Venn diagram for events $A$, $B$, and $C$.

12. Recall that in a prospective cohort study, exposure ($E^+$ or $E^-$) is given, so that the odds ratio is defined as

\[
OR = \frac{\text{odds of disease, given exposure}}{\text{odds of disease, given no exposure}} = \frac{P(D+ \mid E^+) \cdot P(E^- \mid D^+)}{P(D+ \mid E^-) \cdot P(E^+ \mid D^-)}
\]

Recall that in a retrospective case-control study, disease status ($D^+$ or $D^-$) is given; in this case, the corresponding odds ratio is defined as

\[
OR = \frac{\text{odds of exposure, given disease}}{\text{odds of exposure, given no disease}} = \frac{P(E^+ \mid D^+) \cdot P(D^- \mid E^+)}{P(E^+ \mid D^-) \cdot P(D^+ \mid E^-)}
\]

Show algebraically that these two definitions are mathematically equivalent, so that the same “cross product ratio” calculation can be used in either a cohort or case-control study, as the following two problems demonstrate. (Recall the definition of conditional probability.)

13. Suppose that some entries of the joint probability table of two independent events (given by, respectively, the rows and columns below) have been accidentally deleted.

\[
\begin{array}{c|c}
.10 & .20 \\
\hline
.25 & \end{array}
\]

Restore the missing values. (Hint: Start by letting $x$ and $y$ be the first and third column marginal probabilities respectively, then solve for the two row marginal probabilities.)

**NOTE:** There are two possible solutions. Find both of them.
14. Under construction…

15. An observational study investigates the connection between aspirin use and three vascular conditions – gastrointestinal bleeding, primary stroke, and cardiovascular disease – using a group of patients exhibiting these disjoint conditions with the following prior probabilities: 

- \( P(\text{GI bleeding}) = 0.2 \)
- \( P(\text{Stroke}) = 0.3 \)
- \( P(\text{CVD}) = 0.5 \)

as well as with the following conditional probabilities:

- \( P(\text{Aspirin} | \text{GI bleeding}) = 0.09 \)
- \( P(\text{Aspirin} | \text{Stroke}) = 0.04 \)
- \( P(\text{Aspirin} | \text{CVD}) = 0.02 \)

(a) Calculate the following posterior probabilities: 

- \( P(\text{GI bleeding} | \text{Aspirin}) \)
- \( P(\text{Stroke} | \text{Aspirin}) \)
- \( P(\text{CVD} | \text{Aspirin}) \)

(b) Interpret: Compare the prior probability of each category with its corresponding posterior probability. What conclusions can you draw? Be as specific as possible.

16. On the basis of a retrospective study, it is determined (from hospital records, tumor registries, and death certificates) that the overall five-year survival (event \( S \)) of a particular form of cancer in a population has a prior probability of \( P(S) = 0.4 \). Furthermore, the conditional probability of having received a certain treatment (event \( T \)) among the survivors is given by \( P(T | S) = 0.8 \), while the conditional probability of treatment among the non-survivors is only \( P(T | S^c) = 0.3 \).

(a) A cancer patient is uncertain about whether or not to undergo this treatment, and consults with her oncologist, who is familiar with this study. Compare the prior probability of overall survival given above with each of the following posterior probabilities, and interpret in context.

- Survival among treated individuals, \( P(S | T) \)
- Survival among untreated individuals, \( P(S | T^c) \)

(b) Calculate the Relative Risk of survival for this disease. Interpret this value.

(c) Also calculate the following.

- Odds of survival, given treatment
- Odds of survival, given no treatment
- Odds Ratio of survival for this disease. Interpret this value.

17. WARNING! This problem is not for the mathematically timid.

Recall that two events \( A \) and \( B \) are statistically independent if \( P(A \cap B) = P(A) P(B) \). It therefore follows that the difference

\[
\Delta = \left| P(A \cap B) - P(A) P(B) \right|
\]

is a measure of “how far” from statistical independence any two arbitrary events \( A \) and \( B \) are. Prove that \( \Delta \leq \frac{1}{4} \). When is the inequality sharp? (That is, when is equality achieved?)
18. First, recall that, for any two events $A$ and $B$, the union $A \cup B$ defines the “inclusive or” – i.e., “Either $A$ occurs, or $B$ occurs, or both.”

Now, consider the event “Only $A$” – i.e., “Event $A$ occurs, and event $B$ does not occur” – defined as the intersection $A \cap B^c$, also denoted as the difference $A - B$. Likewise, “Only $B$” = “$B$ and not $A$” = $B \cap A^c = B - A$. Using these, we can define “$\text{xor}$” – the so-called “exclusive or” – i.e., “Either $A$ occurs, or $B$ occurs, but not both” – as the union $(A - B) \cup (B - A)$, or equivalently, $(A \cup B) - (A \cap B)$. This is also sometimes referred to the symmetric difference between $A$ and $B$, denoted $A \Delta B$. (See the two regions corresponding to the highlighted formulas below.)

(a) Suppose that two treatment regimens $A$ and $B$ exist for a certain medical condition. It is reported that 35% of the total patient population receives Treatment $A$, 40% receives Treatment $B$, and 14% receives both treatments. Construct the corresponding Venn diagram and $2 \times 2$ probability table. Are the two treatments $A$ and $B$ statistically independent of one another?

Calculate $P(A \text{ or } B)$, and $P(A \text{ xor } B)$.

(b) Suppose it is discovered that an error was made in the original medical report, and it is actually the case that 35% of the population receives only Treatment $A$, 40% receives only Treatment $B$, and 14% receives both treatments. Construct the corresponding Venn diagram and $2 \times 2$ probability table. Are the two treatments $A$ and $B$ statistically independent of one another?

Calculate $P(A \text{ or } B)$, and $P(A \text{ xor } B)$.
19. Three of the most common demographic variables used in epidemiological studies are age, sex, and race. Suppose it is known that, in a certain population,

- 30% of whites are men, 40% of males are white men, 50% of white males are men.

(a) What percentage of whites are male? **Formally justify your answer!**

(b) What percentage of males are white? **Formally justify your answer!**

*Hint:* Follow the same notation as the example in section 3.2, slide 26, of the PowerPoint slides.

20. In another epidemiological study, it is known that, for a certain population,

- 10% of adults are men, 20% of males are white, 30% of whites are adults
- 40% of males are men, 50% of whites are male.

What percentage of adults are white?

*Hint:* Find a connection between the products \( P(A | B) P(B | C) P(C | A) \) and \( P(B | A) P(C | B) P(A | C) \).

21. **The Shell Game.** In the traditional version, a single pea is placed under one of three walnut half-shells in full view of an observer. The shells are then quickly shuffled into a new random arrangement, and the observer then guesses which shell contains the pea. If the guess is correct, the observer wins.

(a) For the sake of argument, suppose there are 20 half-shells instead of three, and the observer plays the game a total of \( n \) times. What is the probability that he/she will guess correctly at least once out of those \( n \) times? How large must \( n \) be, in order to guarantee that the probability of winning is over 50%? What happens to the probability as \( n \to \infty \)?

(b) Now suppose there are \( n \) half-shells, and the observer plays the game a total of \( n \) times. What is the probability that he/she will guess correctly at least once out of those \( n \) times? What happens to this probability as \( n \to \infty \)?

*Hint (for both parts):* First calculate the probability of losing all \( n \) times.

22. (a) By definition, two events \( A \) and \( B \) are **statistically independent** if and only if \( P(A | B) = P(A) \). Prove mathematically that two events \( A \) and \( B \) are independent if and only if \( P(A | B) = P(A | B^c) \).

*Hint: Let \( P(A) = a, \ P(B) = b, \ P(A \cap B) = c, \) and use either a Venn diagram or a \( 2 \times 2 \) table.*

(b) More generally, let events \( A, B_1, B_2, \ldots, B_n \) be defined as in Bayes’ Theorem. Prove that:

\( A \) and \( B_1 \) are independent, \( A \) and \( B_2 \) are independent, \ldots, \( A \) and \( B_n \) are independent

if and only if \( P(A | B_1) = P(A | B_2) = \ldots = P(A | B_n) \).

*Hint: Use the Law of Total Probability.*
23. Prove that the relative risk $RR$ is always between 1 and the odds ratio $OR$. (Note there are three possible cases to consider: $RR < 1$, $RR = 1$, and $RR > 1$.)

24. Consider the following experiment. Pick a random integer from 1 to $10^{12}$.
   (a) What is the probability that it is either a perfect square (1, 4, 9, 16, ...) or a perfect cube (1, 8, 27, 64, ...)?
   (b) What is the probability that it is either a perfect fourth power (1, 16, 81, 256, ...) or a perfect sixth power (1, 64, 729, 4096, ...)?

25. As defined at the beginning of this chapter, the probability of Heads of a coin is formally identified with $\lim_{n \to \infty} \frac{X(n)}{n}$ – when that limiting value exists – where $n =$ # tosses, and $X =$ # Heads in those $n$ tosses. Show by a mathematical counterexample that in fact, this limit need not necessarily exist. That is, provide an explicit sequence of Heads and Tails (or ones and zeros) for which the ratio $\frac{X(n)}{n}$ does not converge to a unique finite value, as $n$ increases.

26. Warning: These may not be quite as simple as they look.
   (a) Consider two independent events $A$ and $B$. Suppose $A$ occurs with probability 60%, while “B only” occurs with probability 30%. Calculate the probability that $B$ occurs, i.e., $P(B)$.
   (b) Consider two independent events $C$ and $D$. Suppose they both occur together with probability 72%, while there is a 2% probability that neither event occurs. Calculate the probabilities $P(C)$ and $P(D)$.

27. Solve for the middle cell probability (“?”) in the following partially-filled probability table.

<table>
<thead>
<tr>
<th></th>
<th>.01</th>
<th>?</th>
<th>.02</th>
</tr>
</thead>
<tbody>
<tr>
<td>.03</td>
<td>?</td>
<td>.04</td>
<td>.50</td>
</tr>
<tr>
<td>.60</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

28. How far away can a prior probability be from its posterior probabilities? Consider two events $A$ and $B$, and let $P(A \mid B) = p$ and $P(A \mid B^c) = q$ be fixed probabilities. If $p = q$, then $A$ and $B$ are statistically independent (see problem 22 above), and thus the prior probability $P(B)$ coincides with its corresponding posterior probabilities $P(B \mid A)$ and $P(B \mid A^c)$ exactly, yielding a minimum value of 0 for the absolute differences $|P(B) - P(B \mid A)|$ and $|P(B) - P(B \mid A^c)|$.

In terms of $p$ and $q$ (with $p \neq q$), what must $P(B)$ be for the maximum absolute differences to occur, and what are their respective values?
29. Let $A$, $B$, and $C$ be three pairwise-independent events, that is, $A$ and $B$ are independent, $B$ and $C$ are independent, and $A$ and $C$ are independent. It does not necessarily follow that $P(A \cap B \cap C) = P(A)P(B)P(C)$, as the following Venn diagram illustrates. Provide the details.

\[ 1 - a - b - c + ab + ac + bc - d \]

30. **Bar Bet**

(a) Suppose I ask you to pick any four cards at random from a deck of 52, without replacement, and bet you one dollar that at least one of the four is a face card (i.e., Jack, Queen, or King). Should you take the bet? Why? *(Hint: See how the probability of this event compares to 50%. If this is too hard, try it with replacement first.)*

(b) What if the bet involves picking three cards at random instead of four? Should you take the bet then? Why?

(c) Refer to the posted Rcode folder for this part. **Please answer all questions.**
31.
(a) True or False? “If event $A$ and event $B$ are statistically independent, and if event $A$ and event $C$ are statistically independent, then it follows that event $B$ and event $C$ are statistically independent.” Prove or find a counterexample (e.g., via a Venn diagram).

(b) Repeat part (a), with the word “independent” replaced by “dependent.”

32.
(a) True or False? “If event $A$ and event $B$ are statistically independent, and if event $A$ and event $C$ are statistically independent, then it follows that event $A$ and event $B \cap C$ – i.e., ‘$B$ and $C$’ – are statistically independent.” Prove or find a nontrivial counterexample (e.g., via a Venn diagram).

(b) True or False? “If event $A$ and event $B$ are statistically independent, and if event $A$ and event $C$ are statistically independent, and if event $A$ and event $B \cap C$ – i.e., ‘$B$ and $C$’ – are statistically independent, then it follows that event $A$ and event $B \cup C$ – i.e., ‘$B$ or $C$’ – are statistically independent.” Prove or find a nontrivial counterexample (e.g., via a Venn diagram).

(c) Repeat part (b), but with the underlined condition replaced by the condition $B \cap C = \emptyset$. (See Venn diagram.)