Recitation 3: Conditional Probability

1 Review

1. $P(A^c) = 1 - P(A)$
2. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
3. Given two events, A and B, the definition of *conditional probability*:
$$P(A|B) = \frac{P(A \cap B)}{P(B)},$$
defined only when $P(B) > 0$.
4. Rearranging terms to obtain the *multiplication rule*:
$$P(A \cap B) = P(A|B)P(B) = P(B|A)P(A)$$
5. Given events $A_1, \ldots, A_n$ that partition the sample space Ω, the *law of total probability*:
$$P(B) = \sum_{i=1}^{n} P(B \cap A_i) = \sum_{i=1}^{n} P(B|A_i)P(A_i)$$
(Partitioning the probability space means that the events do not intersect. i.e. $P(A_i \cap A_j) = 0$ for $i \neq j$.)
6. Combining the above three facts yields *Bayes' Theorem*:
$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_{j=1}^{n} P(B|A_j)P(A_j)}$$
7. Two events A and B are *independent* if
$$P(A \cap B) = P(A)P(B)$$
If $P(B) > 0$, then independence implies $P(A|B) = P(A)$.

2 R-facts

1. In the last recitation we learned the R functions `choose`(n, k) and `factorial`(n).
2. Writing a for-loop in R:

```
for (i in 2010:2015){
    print(paste("The year is", i))
}
```

The output is

```
"The year is 2010"
"The year is 2011"
"The year is 2012"
"The year is 2013"
"The year is 2014"
"The year is 2015"
```

3. You can define a function in R. The function can call itself; this is called a recursion. For example, we can compute factorials recursively by using the fact that $n! = n \cdot (n-1)!$.

```
recur_factorial <- function(n) {
    if(n <= 1) {
        return(1)
    } else {
        return(n * recur_factorial(n-1))
    }
}
```

   Output:

```
> recur_factorial(4)
[1] 24
```

   Bonus: Can you write a recursive function yourself to compute $\binom{n}{k}$?

3 Exercises

1. Suppose we have two fair four-sided dice: one green, one red. Suppose we define the following events:
   - A : The red die shows a 3.
   - B : The green die shows a 3.
   - C : Both dice show the same number.
   - D : The sum of the dice is 7.
   - E : The sum of the dice is even.

   (a) Give two pairs of events from the above list that are mutually exclusive.
   We have a two choices: C and D are mutually exclusive, while D and E are also mutually exclusive.

   (b) Compute $P(A \cap E)$ and $P(A \cup E)$.
   Denote by G_i, $i = 1, \dots, 4$, the outcome of rolling face i with the green die, so the possible outcomes for the green die are $\{G_1, G_2, G_3, G_4\}$. Similarly say the possible outcomes for the red die are $\{R_1, \dots, R_4\}$.
   Notice that $A \cap E = \{(G_1, R_3), (G_3, R_3)\}$ so that $P(A \cap E) = 2/16 = 1/8$. In the same spirit,

$$A \cup E = \{(G_1, R_3), (G_2, R_3), (G_3, R_3), (G_4, R_3), (G_1, R_1),$$
$$(G_2, R_2), (G_2, R_4), (G_3, R_1), (G_4, R_2), (G_4, R_4)\},$$

hence $P(A \cup E) = 10/16 = 5/8$. For $P(A \cup E)$ notice that we could have also used the formula $P(A \cup E) = P(A) + P(E) - P(A \cap E)$. Doing this we get

$$P(A \cup E) = \frac{1}{4} + \frac{1}{2} - \frac{1}{8} = \frac{2+4-1}{8} = \frac{5}{8}.$$

2. In 2005, a survey on drinking habits was conducted among Cornell university students. The survey saw 448 respondents. Of those surveyed, 73% admitted to having consumed alcoholic drinks on at least one occasion within the past 30 days. Of these, 10.7% admitted to having caused their Blood Alcohol Content (BAC) to rise above the legal limit of 0.08 on the first of their drinking occasions from within the past 30 days.

   A person's BAC is commonly measured from their breath using a device known as a Breathalyzer. However, the readings are not always accurate. A Breathalyzer has a false positive rate of 23% (this is the proportion of people whose BAC is below 0.08 but whose reading falls above 0.08), and a false negative rate of 15% (the proportion of people whose BAC is above the legal limit but whose reading falls below).

   (a) Suppose that every respondent who consumed at least one alcoholic drink had their BAC measured. How many would receive a reading falling above the legal limit?
   $0.73 \cdot 448 = 327$ students had at least one drink and $0.107 \cdot 327 = 35$ of them had BAC above the limit. Therefore
   #high readings = (1 − false negative rate) · #above limit + (false positive rate) · (#at least 1 drink − #above limit) = $0.85 \cdot 35 + 0.23 \cdot 292 = 97$

   (b) Assume that the proportions discovered in the survey hold for the entire Cornell community. Now suppose President Pollack received a reading above the legal limit. What is the probability that her BAC is actually above 0.08?
   Let hR-high reading, lR-low reading, hB-high BAC, lB-low BAC. Then from Bayes theorem

   $$P(\text{high BAC}|\text{high reading}) = P(hB|hR) = \frac{P(hR|hB)P(hB)}{P(hR|hB)P(hB) + P(hR|lB)P(lB)}$$
   $$P(hR|hB) = 1 - P(lR|hB) = 1 - 0.15 = 0.85$$
   $$P(hR|lB) = 0.23$$
   $$P(hB) = P(hB|\text{drink})P(\text{drink}) + P(hB|\text{no drink})P(\text{no drink}) = 0.107 \cdot 0.73 + 0 \cdot 0.27 = 0.078$$
   $$P(lB) = 1 - P(hB) = 1 - 0.078 = 0.922$$

   Therefore

   $$P(\text{high BAC}|\text{high reading}) = \frac{0.85 \cdot 0.078}{0.85 \cdot 0.078 + 0.23 \cdot 0.922} = 0.238.$$

3. Suppose you shuffle a standard deck of playing cards, then deal two cards off the top of the deck. Use the Law of Total Probability to compute the probability that you obtain a jack and a queen, in either order. (Remember that a standard deck contains 52 cards, 4 of which are jacks, and 4 of which are queens.)

Let A be the desired event, J and Q be the events where the first card revealed is a jack or queen (respectively), and N be the event where the first card is neither a jack nor a queen. Since J, Q, and N partition the sample space, the Law of Total Probability gives:

$$\begin{aligned} P(A) &= P(A \mid J)P(J) + P(A \mid Q)P(Q) + P(A \mid N)P(N) \\ &= \frac{4}{51} \cdot \frac{4}{52} + \frac{4}{51} \cdot \frac{4}{52} + 0 \cdot \frac{44}{52} \\ &= 0.011. \end{aligned}$$

4. Suppose you roll a fair eight-sided die. Define the events $A = \{1, 3, 5, 8\}$, $B = \{2, 4, 7, 8\}$, and $C = \{3, 5, 6, 8\}$. Show that $P(A \cap B \cap C) = P(A)P(B)P(C)$, but that any two of A, B, and C are not independent.

   Since $P(A) = P(B) = P(C) = 1/2$, and $P(A \cap B \cap C) = P(\{8\}) = 1/8$, we indeed have

$$P(A \cap B \cap C) = P(A)P(B)P(C).$$

   However,

$$\begin{aligned} P(A \cap B) &= P(\{8\}) \neq \frac{1}{4} = P(A)P(B) \\ P(A \cap C) &= P(\{3, 5, 8\}) \neq \frac{1}{4} = P(A)P(C) \\ P(A \cap B) &= P(\{8\}) \neq \frac{1}{4} = P(B)P(C), \end{aligned}$$

   and so pairs of events are dependent.

5. New York City subway trains make around 8000 trips on an average weekday, and there are close to 600 trains operational during the rush hours of 7:30-8:30am and 5:00-6:00pm. During rush hour, trains have a 65% chance of running on time, and it is nearly impossible to find a seat. At all other times, trains are running on time 92% of the time, and it is easy to find a seat. To simplify the problem, think of rush hour as accounting for 1/12 of the day, and assume that during rush hour it is always difficult to find a seat, and outside of rush hour it is always easy to find a seat.

   (a) Your train is running late. What is the probability that it will be difficult to find a seat?

   From Bayes Theorem:
   P(seat difficult | late) = P(rush hr | late) = P(late | rush hr)P(rush hr) / [P(late | rush hr)P(rush hr) + P(late | not rush hr)P(not rush hr)] = 0.35 · (1/12) / [0.35 · (1/12) + 0.08 · (11/12)] = 0.285

   (b) If you ask the average subway customer about their experience, most will likely complain that the trains are always running late. However, according to the given numbers, trains are mostly on time, most of the time. Indeed, the probability that a customer will experience a late train (the denominator of the fraction in the previous question) is approximately 0.1, or 10%. How would you explain this discrepancy?

   Most customers travel during rush hour most of the time, so they would tend to experience late trains more often than customers travelling at any (random) time, who experience 10% lateness.

6. An estimated 20 million Americans (1 in 15) suffer from asthma. Also, an estimated 50 million Americans (1 in 6) are reported to have allergies. Furthermore, it has been determined that 50% of asthma cases are extrinsic (where a person has allergies, and the allergies cause them to develop asthma).

   As an (American) freshman at Cornell, you have a 12% chance of getting placed in one of either Court-Kay-Bauer Hall or Mews Hall. If you suffer from extrinsic asthma, you have a 25% chance of getting placed in one of these.

   Given that you have been placed in either Court-Kay-Bauer Hall or Mews Hall, what is the probability that you suffer from extrinsic asthma?

   P(extrinsic asthma) = 1/2*P(asthma) = 1/30
   P(extr | hall) = P(hall | extr)P(extr)/P(hall) = $0.25 \cdot (1/30)\ /\ 0.12 = 0.069$