Problem Set 2

Problem 1:

Say in Silicon Valley, 35% of engineers program in Java and 28% of the engineers who program in Java also program in C++. Furthermore, 40% of engineers program in C++.
(a) What is the probability that a randomly selected engineer programs in Java and C++?
(b) What is the conditional probability that a randomly selected engineer programs in Java given that they program in C++?

Solution:

(a) **0.098**. $P(\text{Java}, \text{C++}) = P(\text{Java}) * P(\text{C++}|\text{Java}) = 0.35 * 0.28 = 0.098.$
(b) **0.245**. We are asked to find $P(J|C++)$, which is $\frac{P(J,C++)}{P(C++)}$, Since we know the both of them happening is 0.098 and the probability of someone knowing C++ is 0.40, $P(J|C++) = \frac{0.098}{0.40} = 0.245.$

Problem 2:

A website wants to detect if a visitor is a robot or a human. They give the visitor five CAPTCHA tests that are hard for robots but easy for humans. If the visitor fails one of the tests, they are flagged as a robot. The probability that a human succeeds at a single test is 0.95, while a robot only succeeds with probability 0.3. Assume all tests are independent. The percentage of visitors on this website that are robots is 5%; all other visitors are human.
(a) If a visitor is actually a robot, what is the probability they get flagged (the probability they fail at least one test)?
(b) If a visitor is human, what is the probability they get flagged?
(c) Suppose a visitor gets flagged. Using your answers from part (a) and (b), what is the probability that the visitor is a robot?
(d) If a visitor is human, what is the probability that they pass exactly three of the five tests?
(e) Building off of your answer from part (d), what is the probability that a visitor with unknown identity passes exactly three of the five tests?

Solution:

(a) **0.99757**. With a 0.3 probability of passing each independent test the probability of a robot passing all the tests are 0.3^5, as for the probability of a robot getting flagged is $1 - 0.3^5 = 0.99757$.
(b) **~0.226**. Similarly to part (a) the probability of passing each test is 0.95^5, as for probability of getting flagged is $1 - 0.95^5 = 0.2262190625$.
(c) **~0.1884**. We want the conditional probability that the probability of the visitor being a robot given that he is flagged. We can find that by breaking the conditional probability into the probability of things that we know:

$$P(R|F) = \frac{P(R,F)}{P(F)} = \frac{P(F|R)P(R)}{P(F|R)P(R)+P(F|R^C)P(R^C)}$$
$$\text{we know the nominator which is: } 0.99757 * 0.05 = 0.0498785$$
$$\text{we know the denominator which is: } 0.99757 * 0.05 + 0.2262190625 * 0.95 = 0.2647866094$$

Using equation and the values we get the answer: $P(R|F) = \frac{0.0498785}{0.2647866094} = 0.1883724412$.
(d) **~0.0214**. Any visitor whether they are a robot or a human passing exactly 3 tests and failing 2 tests is: $\binom{5}{3} * p^3 * (1-p)^2$ given that p is the probability of passing one test. A human has a p value of 0.95 so answer is: $\binom{5}{3} * (0.95)^3 * (0.05)^2 = 0.021434375$.
(e) **~0.02698**. Probability of a visitor with unknown identity passing exactly three tests and failing two tests is the sum of the the probability of passing each three tests and failing 2 tests for both a human and a robot: $\binom{5}{3} * (0.3)^3 * (0.7)^2 = 0.1323$ is the probability for a robot and since we know the probability for a human we can just find the answer: $0.1323 * 0.05 + 0.021434375 * 0.95 = 0.0269776563$.

Problem 3:

Say all computers either run operating system W or X. A computer running operating system W is twice as likely to get infected with a virus as a computer running operating system X. If 70% of all computers are running operating system W, what percentage of computers infected with a virus are running operating system W?

Solution:

~82.35%. Let p be the chance of someone using the system X getting infected with a virus then the probability of someone using the system W is 2p. Since 0.70 of all computers use system W the amount of computers that is infected with a virus can be found by, $2p * 0.70$. Similarly the amount of infected computers using system x is $p * 0.30$. The total amount of computers affected by the virus is $0.30p + 1.40p = 1.70p$. Now we can find the probability of the infected computer using the system W:
$\frac{1.4p}{1.7p} = 0.8235294118$.

Problem 4:

The Superbowl institutes a new way to determine which team receives the kickoff first. The referee chooses with equal probability one of three coins. Although the coins look identical, they have probability of heads 0.1, 0.5 and 0.9, respectively. Then the referee tosses the chosen coin 3 times. If more than half the tosses come up heads, one team will kick off; otherwise, the other team will kick off. If the tosses resulted in the sequence H, T, H, what is the probability that the fair coin was actually used?

Solution:

~0.5814. $0.1^2 * 0.9$, $0.5^2 * 0.5$, and $0.9^2 * 0.1$ are the probabilities of getting H, T, H on the three coin flips respectively. We are trying to find conditional probability $P(A|X)$, for A which is the fair coin and X which is the H, T, H sequence. To find the conditional probability we do the same thing we did with part (c) of problem 2:

$$P(A|X) = \frac{P(A,X)}{P(X)} = \frac{P(X|A)P(A)}{P(X,A)+P(X,A^C)} = \frac{P(X|A)P(A)}{P(X|A)P(A)+P(X|A^C)P(A^C)}$$
$$\text{we know the nominator which is: } 0.125 * \frac{1}{3} = 0.0416666667$$
$$P(X|A^C)P(A^C) = 0.1^2 * 0.9 * \frac{1}{3} + 0.9^2 * 0.1 * \frac{1}{3} = 0.003 + 0.027 = 0.03$$
$$\text{we know the denominator which is: } 0.0416666667 + 0.03 = 0.0716666667$$

Therefore the answer is: $\frac{0.0416666667}{0.0716666667} = 0.581395349$.

Problem 5:

After a long night of programming, you have built a powerful, but slightly buggy, email spam filter. When you don't encounter the bug, the filter works very well, always marking a spam email as SPAM and always marking a non-spam email as GOOD. Unfortunately, your code contains a bug that is encountered 10% of the time when the filter is run on an email. When the bug is encountered, the filter always marks the email as GOOD. As a result, emails that are actually spam will be erroneously marked as GOOD when the bug is encountered. Let p denote the probability that an email is actually non-spam, and let q denote the conditional probability that an email is non-spam given that it is marked as GOOD by the filter.
(a) Determine q in terms of p.
(b) Using your answer from part (a), explain mathematically whether q or p is greater. Also, provide an intuitive justification for your answer.

Solution:

(a) $\boldsymbol{q = \frac{p}{0.9p+0.1}}$. $q = P(ns|good)$ and $p = P(ns)$ to find q in terms of p we can separate q into smaller steps: $P(ns|good) = \frac{P(ns,good)}{P(good)}$. We can't place p anywhere so we continue: $\frac{P(good|ns)P(ns)}{P(good)}$, we can find the nominator for this expression which is just p since all the non-spam email gets tagged good there is no need to calculate it but we need to break

the denominator into smaller steps: $\frac{p}{P(good,ns)+P(good,ns^C)}$, now we can also calculate for the denominator which is: $P(good, ns) = p$ like the nominator; $P(good, ns^C) = 0.1 * (1-p)$. So the final answer is $q = \frac{p}{0.9p+0.1}$.
(b) q **>** p. q is bigger because $0.9p$ will always be smaller than 0.1 as long as $0 < p < 1$ therefore q will be bigger only when p is 1 or 0 will q be equal to p.

Problem 6:

Two cards are randomly chosen without replacement from an ordinary deck of 52 cards. Let E be the event that both cards are Aces. Let F be the event that the Ace of Spades is one of the chosen cards, and let G be the event that at least one Ace is chosen.
(a) Compute P(E | F).
(b) Are E and F independent? Justify your answer using your response to part (a).
(c) Compute P(E | G).

Solution:

(a) **~0.05882**. we can find this using combinatorics: $P(E|F) = \frac{\binom{3}{1}\binom{1}{1}}{\binom{51}{1}} = 0.0588235294$, nominator $\binom{3}{1}\binom{1}{1}$ since we have one of the aces we just need to choose one ace from the remaining three suits.
(b) **Not independent**. They are not independent because the in the event E we are choosing 2 aces from 4 suits which is 6 different choices but when we condition E on the event F the n the choices goes down to 3 since we are choosing 1 ace from 3 suits making the probabilities different.
(c) **~0.0303**. we can find this using combinatorics: $P(E|G) = \frac{\binom{4}{2}}{\binom{4}{2}+\binom{4}{1}\binom{48}{1}} = 0.0303030303$, nominator is $\binom{4}{2}$ since we need to pick 2 aces from two of the 4 suits; denominator is $\binom{4}{2} + \binom{4}{1}\binom{48}{1}$ since for G we have two choices one is that we choose both cards ace and the other one is we choose one ace and the other some other card.

Problem 7:

Your colleagues in a comp-bio lab have sequenced DNA from a large population in order to understand how a gene (G) influences two particular traits (T1 and T2). They find that P(G) = 0.6, P(T1 | G) = 0.7, and P(T2 | G) = 0.9. They also observe that if a subject does not have the gene G, they express neither T1 nor T2. The probability of a patient having both T1 and T2 given that they have the gene G is 0.63.
(a) Are T1 and T2 conditionally independent given G?
(b) Are T1 and T2 conditionally independent given G^C?
(c) What is P(T1)?
(d) What is P(T2)?
(e) Are T1 and T2 independent?

Solution:

We know:
$P(G) = 0.6, \quad P(T_1|G) = 0.7, \quad P(T_2|G) = 0.9, \text{ and } P(T_1, T_2|G) = 0.63$
(a) **Yes**. Since $P(T_1, T_2|G) = P(T_1|G) * P(T_2|G) = 0.63$.
(b) **Yes**. Since the T_1 and T_2 traits are not expressed when the subject does not have the gene G: $P(T_1, T_2|G^C) = P(T_1|G^C) * P(T_2|G^C) = 0$.
(c) **0.42**. We can calculate $P(T_1)$ by separating $P(T_1|G)$ until we see $P(T_1)$:
$P(T_1|G) = \frac{P(T_1,G)}{P(G)} = \frac{P(G|T_1)P(T_1)}{P(G)} \quad \rightarrow \quad 0.7 = \frac{P(G|T_1)P(T_1)}{0.6} \rightarrow P(T_1) = 0.42$(Notice that $P(G|T_1)$ is 1 since without G there is no T_1)
(d) **0.54**. Calculation for T_2 is same as part (c).
(e) **Not independent**. For T_1 and T_2 to independent the following expression must be true: $P(T_1, T_2) = P(T_1)P(T_2)$. We can find $P(T_1, T_2)$ by:
$P(T_1, T_2) = P(T_1, T_2|G)P(G) + P(T_1, T_2|G^C)P(G^C) \quad \rightarrow \quad P(T_1, T_2) = P(T_1, T_2|G)P(G) = 0.63 *$
. Now we calculate $P(T_1)P(T_2) = 0.42 * 0.54 = 0.2268$. Since $0.378 \neq 0.2268$ T_1 and T_2 are not independent.

Problem 8:

The color of a person's eyes is determined by a pair of eye-color genes, as follows:

- if both of the eye-color genes are blue-eyed genes, then the person will have blue eyes
- if one or more of the genes is a brown-eyed gene, then the person will have brown eyes
  A newborn child independently receives one eye-color gene from each of its parents, and the gene it receives from a parent is equally likely to be either of the two eye-color genes of that parent. Suppose William and both of his parents have brown eyes, but William's sister (Claire) has blue eyes. (We assume that blue and brown are the only eye-color genes.)
  (a) What is the probability that William possesses a blue-eyed gene?
  (b) Suppose that William's wife has blue eyes. What is the probability that their first child will have blue eyes?

Solution:

- Possible eye color genes: {BB, Bb, bB(dominant gene is always shown ahead of the recessive genes but for this question I will show it in this way), bb} where B is brown-eyed gene and b is blue-eyed gene.
- William's possible eye genes: {BB, Bb, bB}.
- Parents' possible eye genes: {Bb} since one of the child's eye color is blue than both parents must have exactly one blue-eyed gene and one brown-eyed gene.

- Claire's possible eye genes: {bb}.
- William's wife's possible eye genes: {bb}.
  (a) **~0.6667**. Since There is 2 possible gene combination for this event out of 3 gene combinations the probability is: $\frac{2}{3}$.
  (b) **0.75**. It is the conditional probability of the child having blue eyes given that William also has blue-eyed gene: Since William needs to have a blue-eyed gene for the child to have one $P(W_b|C_b) = 1$ and since the child has a 0.5 chance of getting the blue-eyed gene from William $P(C_b) = \frac{1}{2}$. The answer is
  $P(W_b) * P(C_b) = \frac{2}{3} * \frac{1}{2} = 0.3334.$

Problem 9:

Consider the following algorithm for betting in [roulette][https://en.wikipedia.org/wiki/Roulette]. At each round ("spin"), you bet $1 on a color ("red" or "black"). If that color comes up on the wheel, you keep your bet AND win $1; otherwise, you lose your bet.

- i. Bet $1 on "red"
- ii. If "red" comes up on the wheel (with probability 18/38), then you win $1 (and keep your original $1 bet) and you immediately quit (i.e., you do not do step (iii) below).
- iii. If "red" did not come up on the wheel (with probability 20/38), then you lose your initial
  $1 bet. But, then you bet \$1 on ``red'' on each of the next two spins of the wheel. After those two spins,$
  . (Rhetorical question: Would you play this game?)

Solution:

(a) **~0.592**. There are two scenarios which can happen first being we win the first roulette spin and then we which has a probability of $\frac{18}{38}$, next one is we lose the first bet $\frac{20}{38}$ and bet on the next two spins and win both of them two get $X > 0$, $(\frac{18}{38})^2$. So
$P(X > 0) = \frac{18}{38} + \frac{20}{38} * ((\frac{18}{38})^2) = 0.591777227.$
(b) **~0.108**.
$E[X] = [1 * \frac{18}{38}] + [(-3) * (\frac{20}{38})^3] + [(-1) * 2 * \frac{18}{20} * (\frac{20}{38})^2] + [1 * \frac{20}{38} * (\frac{18}{38})^2] = -0.108033241$
.