Homework 5

CS 70, Summer 2024

1 Free Points

Make sure you filled out the form!

2 Disease Testing

Let $+$ be the event that the test returns a positive, $-$ be the event that the test returns a negative, D be the event that the person has the flue, and D^{C} be the event that the person does not have the flu.

By the question statement, we have the following probabilities.

$$\mathrm{P}(+ \mid D) = 0.9, \qquad \mathrm{P}(- \mid D^{\mathrm{C}}) = 0.95, \qquad \mathrm{P}(D) = 0.04.$$

(a) We need to find $\mathrm{P}(D \mid +)$. Since this is a conditional probability, we'll need to find it using either Bayes rule or division rule. We notice that we're trying to reverse the direction of conditioning: we know $\mathrm{P}(+ \mid D)$ and are trying to find $\mathrm{P}(D \mid +)$. That's our cue to use Bayes rule.

So we get that

$$\mathrm{P}(D \mid +) = \frac{\mathrm{P}(+ \mid D)\mathrm{P}(D)}{\mathrm{P}(+)} = \frac{0.90 \cdot 0.04}{\mathrm{P}(+)}.$$

To find the denominator, note that we can find the chance of a positive result if we know whether the person has the disease or not. That's our cue to partition (also known as the law of total probability). We have that

$$\begin{aligned}\mathrm{P}(+) &= \mathrm{P}(+ \cap D) + \mathrm{P}(+ \cap D^{\mathrm{C}}) \\ &= \mathrm{P}(+ \mid D)\mathrm{P}(D) + \mathrm{P}(+ \mid D^{\mathrm{C}})\mathrm{P}(D^{\mathrm{C}}) \\ &= \mathrm{P}(+ \mid D)\mathrm{P}(D) + \big(1 - \mathrm{P}(- \mid D^{\mathrm{C}})\big)(1 - \mathrm{P}(D)) \\ &= 0.90 \cdot 0.04 + (1 - 0.95) \cdot (1 - 0.04).\end{aligned}$$

So the probability is

$$\mathrm{P}(D \mid +) = \frac{0.90 \cdot 0.04}{0.90 \cdot 0.04 + (1 - 0.95) \cdot (1 - 0.04)} \approx 43\%.$$

Notice that this is quite low despite the high conditional probabilities of the test being correct given the person's disease status. This is because so few people in the population (only 4%) have the disease. So we can see that the person is now ten times more likely to have the disease than they were before getting the positive result, but it's still only a 43% chance.

(b) We need to find $\mathrm{P}(D \mid +-)$. We'll use precisely the same approach as in **(a)**, since we're more easily able to find something like $\mathrm{P}(+- \mid D)$. By Bayes rule,

$$\mathrm{P}(D \mid +-) = \frac{\mathrm{P}(+- \mid D)\mathrm{P}(D)}{\mathrm{P}(+-)}.$$

We know that the test results are independent given the person's disease status, so we have that

$$\mathrm{P}(+- \mid D) = \mathrm{P}(+ \mid D)\mathrm{P}(- \mid D) = \mathrm{P}(+ \mid D)(1 - \mathrm{P}(- \mid D) = 0.90 \cdot (1 - 0.90).$$

To find the denominator, we'll leverage the fact that we can find the chance of a positive and a negative test result given the disease status and partition on the disease status. We've already calculated the $\mathrm{P}(+- \mid D)$ part of the first term.

$$\begin{aligned}\mathrm{P}(+-) &= \mathrm{P}(+ - \cap D) + \mathrm{P}(+ - \cap D^{\mathrm{C}}) \\ &= \mathrm{P}(+- \mid D)\mathrm{P}(D) + \mathrm{P}(+- \mid D^{\mathrm{C}})\mathrm{P}(D^{\mathrm{C}}) \\ &= \mathrm{P}(+ \mid D)\mathrm{P}(- \mid D)\mathrm{P}(D) + \mathrm{P}(+ \mid D^{\mathrm{C}})\mathrm{P}(- \mid D^{\mathrm{C}})\mathrm{P}(D^{\mathrm{C}}) \\ &= \mathrm{P}(+ \mid D)\big(1 - \mathrm{P}(+ \mid D)\big)\mathrm{P}(D) + \big(1 - \mathrm{P}(- \mid D^{\mathrm{C}})\big)\mathrm{P}(- \mid D^{\mathrm{C}})\big(1 - \mathrm{P}(D)\big) \\ &= 0.90 \cdot (1 - 0.90) \cdot 0.04 + (1 - 0.95) \cdot 0.95 \cdot (1 - 0.04).\end{aligned}$$

So the probability is

$$P(D \mid +-) = \frac{0.90 \cdot (1-0.90) \cdot 0.04}{0.90 \cdot (1-0.90) \cdot 0.04 + (1-0.95) \cdot 0.95 \cdot (1-0.04)} \approx 7\%$$

We can see that the negative test result brought the chance back down quite a bit: from 43% down to 7%.

(c) Let C be the event that the randomly selected person is a child aged 5 or under. Then we are given that

$$P(D \mid C) = 0.22 \qquad P(D \mid C^{C}) = 0.01.$$

We are being asked to find

$$P(C \mid D) = \frac{P(D \mid C)P(C)}{P(D)}.$$

We know $P(D \mid C)$ and $P(D)$. So we can find the answer if we can find $P(C)$; otherwise we can't.

We know the overall proportion of people who have the flu is 4%, and we know that 22% of children aged 5 or under have the flu and 1% of everyone else has the flue. That 4% is some kind of average of that 22% and 1%. In particular, by partitioning, we have that

$$\begin{aligned} P(D) = 0.04 &= P(D \cap C) + P(D \cap C^{C}) \\ &= P(D \mid C)P(C) + P(D \mid C^{C})P(C^{C}) \\ &= 0.22P(C) + 0.01(1 - P(C)) \\ &= 0.21P(C) + 0.01 \end{aligned}$$

Solving for $P(C)$ yields

$$P(C) = \frac{0.04 - 0.01}{0.21} = \frac{0.03}{0.21} = \frac{1}{7}.$$

Therefore we have that

$$P(C \mid D) = \frac{P(D \mid C)P(C)}{P(D)} = \frac{0.22 \cdot \frac{1}{7}}{0.04} = \frac{22}{28} = \frac{11}{14} \approx 79\%.$$

So 79% of those with the flu are children aged 5 or under.

3 To Bound or Not to Bound

(a) Let S be the event that at least one person gets a suit. This is a union. We could write this event as $S = E_1 \cup E_2 \cup E_3 \cup E_4$, where E_i is the chance that exactly i people get a suit. This will work, but each term requires inclusion-exclusion, since to find the chance that exactly one person gets a suit we need to first make sure they get a suit, and then exclude the overlaps which include other people getting suits, and so on.

A cleaner approach instead defines $S = S_1 \cup S_2 \cup S_3 \cup S_4$ where S_i is the chance that the i^{th} person gets a suit. Then by inclusion-exclusion,

$$P(S) = \sum_{\{i\}} P(S_i) - \sum_{\{i,j\}} P(S_i \cap S_j) + \sum_{\{i,j,k\}} P(S_i \cap S_j \cap S_k) - P(S_1 \cap S_2 \cap S_3 \cap S_4).$$

Note that by symmetry, we have that the summands in each sum all have equal probability: all the singleton probabilities are equal to one another, all the pairwise intersection probabilities are equal to each other, and so on. This is because each person must be equally likely to have a suit (otherwise it's possible to be more likely to get a suit by being dealt first or last, which contradicts the fact that the deck is well-shuffled), each pair is equally likely to have a suit, and so on. So we can simplify the sums as follows:

$$P(S) = \binom{4}{1}P(S_1) - \binom{4}{2}P(S_1 \cap S_2) + \binom{4}{3}P(S_1 \cap S_2 \cap S_3) - \binom{4}{4}P(S_1 \cap S_2 \cap S_3 \cap S_4).$$

To find $P(S_1)$, define our outcome space to be the set of 13-card hands. There are $\binom{52}{13}$ such hands, each of which are equally likely. There is exactly one hand with a full suit of diamonds, so the chance of getting a full suit of diamonds is $1/\binom{52}{13}$. However, the first person could instead have gotten one of the other three suits, so there are a total of 4 hands with a suit. So we have that

$$P(S_1) = 4\frac{1}{\binom{52}{13}}.$$

Now define our outcome space to be the set of two 13-card hands. There are $\binom{52}{13}\binom{39}{13}$ such hands. The chance that the first person gets a full suit of diamonds and the second person gets a full suit of clubs is

$$\frac{1}{\binom{52}{13}\binom{39}{13}}.$$

There are $4 \cdot 3$ such ways for the first person to get all of one suit and the second person to get all of another, so there are

$$P(S_1 \cap S_2) = 4 \cdot 3 \frac{1}{\binom{52}{13}\binom{39}{13}}.$$

Continuing this gets us that

$$P(S_1 \cap S_2 \cap S_3) = 4 \cdot 3 \cdot 2 \frac{1}{\binom{52}{13}\binom{39}{13}\binom{26}{13}} \qquad P(S_1 \cap S_2 \cap S_3 \cap S_4) = 4 \cdot 3 \cdot 2 \cdot 1 \frac{1}{\binom{52}{13}\binom{39}{13}\binom{26}{13}\binom{13}{13}}.$$

Therefore we have that

$$P(S) = \binom{4}{1}\frac{4}{\binom{52}{13}} - \binom{4}{2}\frac{4 \cdot 3}{\binom{52}{13}\binom{39}{13}} + \binom{4}{3}\frac{4 \cdot 3 \cdot 2}{\binom{52}{13}\binom{39}{13}\binom{26}{13}} - \binom{4}{4}\frac{4 \cdot 3 \cdot 2 \cdot 1}{\binom{52}{13}\binom{39}{13}\binom{26}{13}\binom{13}{13}}.$$

(b) This is an intersection. Let A_i be the event that the i^{th} students attends office hours. We want to find $P(A_1 \cap \ldots \cap A_8)$.

Since we don't know whether these events are mutually exclusive or independent, nor do we know any conditional probabilities, none of our rules will work, either for intersections or for unions. So we must bound the chance.

By drawing the events in a Venn diagram, we can see that the intersection is maximized when all the events perfectly overlap. So $P(A_1 \cap \ldots A_8) \leq 0.95$.

To get a lower bound on the intersection, we can consider the complement

$$1 - P(A_1 \cap \ldots \cap A_8) = P(A_1^C \cup \ldots \cup A_8^C),$$

where $P(A_i^C) = 1 - P(A_i) = 0.05$. To bound this chance, draw a Venn diagram to see that the union is maximized when they are all mutually exclusive, in which case the chance of the union is the sum of the probabilities, so

$$P(A_1^C \cup \ldots \cup A_8^C) \leq 8(0.05) = 0.40.$$

Then

$$P(A_1 \cap \ldots \cap A_8) = 1 - P(A_1^C \cup \ldots \cup A_8^C) \geq 1 - 0.40 = 0.60.$$

So we have that

$$0.60 \leq P(A_1 \cap \ldots \cap A_8) \leq 0.95.$$

(c) This is an intersection. Let N_i be the event that the i^{th} bin is not empty. Then the event N that no bin is empty is $N_1 \cap \ldots \cap N_n$.

However, we will quickly get stuck when trying to find

$$P(N_1 \cap \ldots \cap N_n) = P(N_1) \cdot P(N_2 \mid N_1) \cdot \ldots \cdot P(N_n \mid N_1, \ldots, N_{n-1}).$$

In particular, since there are so many ways for the event N_1 to happen, it is very difficult to compute $P(N_2 \mid N_1)$.

So instead, we'll try taking the complement. Let $E_i = N_i^C$ be the event that the i^{th} bin is empty. Then the complement of the event that no bins are empty is the event that at least one bin is empty:

$$P(N^C) = P(E_1 \cup \ldots \cup E_n).$$

These events are not mutually exclusive, so we cannot use addition rule. Instead, we need to use inclusion-exclusion.

We have that

$$P\left(\bigcup_{i=1}^{n} E_i\right) = \sum_{\{i\}} P(E_i) - \sum_{\{i,j\}} P(E_i \cap E_j) + \ldots + (-1)^{n-1} P\left(\bigcap_{i=1}^{n} E_i\right).$$

This is identical to Homework 4, Question 3**(c)** (except with probabilities rather than cardinalities) and Homework 4, Question 6**(d)** (except with n faces rather than 6).

We get by symmetry, since each bin is equally likely to be empty as every other bin, that

$$P\left(\bigcup_{i=1}^{n} E_i\right) = \binom{n}{1}P(E_1) - \binom{n}{2}P(E_i \cap E_j) + \ldots + (-1)^{n-1}\binom{n}{n}P\left(\bigcap_{i=1}^{n} E_i\right).$$

Then by using balls and bins, we get that

$$P(E_1) = \left(1 - \frac{1}{n}\right)^k$$

$$P(E_1 \cap E_2) = \left(1 - \frac{1}{n}\right)^k$$

$$\vdots$$

$$P\left(\bigcap_{i=1}^{n} E_i\right) = \left(1 - \frac{n}{n}\right)^k.$$

Therefore we have that

$$P\left(\bigcup_{i=1}^{n} E_i\right) = \sum_{i=1}^{n}(-1)^{i-1}\binom{n}{i}\left(1 - \frac{i}{n}\right)^k.$$

Then

$$\begin{aligned} P(N) &= 1 - P\left(\bigcup_{i=1}^{n} E_i\right) \\ &= 1 - \sum_{i=1}^{n}(-1)^{i-1}\binom{n}{i}\left(1 - \frac{i}{n}\right)^k \\ &= \sum_{i=0}^{n}(-1)^{i}\binom{n}{i}\left(1 - \frac{i}{n}\right)^k. \end{aligned}$$

(d) Let S_i be the event that there is at least one birthday in the i^{th} season. Observe that this is like throwing n balls into 4 bins labelled with the four seasons and finding the chance that at least one bin is empty. But we just found that in part **(c)** (at the end, right before taking the complement). So the answer is

$$P(S_1 \cup S_2 \cup S_3 \cup S_4) = \sum_{i=1}^{4}(-1)^{i-1}\binom{4}{i}\left(1 - \frac{i}{4}\right)^n.$$

(e) Let R_i be the event that it rains on the i^{th} day of the week. Then the chance of the event R that it rains some day next week is the chance of the union $R_1 \cup \ldots \cup R_7$.

By drawing the Venn diagram, we can see that the chance of the union is minimized when all the events are contained in one another, in which case the union has the same probability as the largest event. So $P(R) \geq 0.15$.

The chance is maximized when all the events are mutually exclusive, in which case the probabilities add. Therefore $P(R) \leq 0.05 + 0.02 + 0.10 + 0.03 + 0.04 + 0.15 + 0.12 = 0.51$.

Therefore

$$0.15 \leq P(R) \leq 0.51.$$

4 Joint Probabilities

(a) If we knew the value of Y, we would be able to find the chance that $X = Y$. In particular, if $Y = k$, then the chance that $X = Y$ is just the chance that X also equals k: p_k.

This is our cue to partition on the value of Y. We get that

$$P(X = Y) = \sum_{k=1}^{N} P(X = Y \cap Y = k)$$

$$= \sum_{k=1}^{N} \mathrm{P}(X = Y \mid Y = k)\mathrm{P}(Y = k)$$

$$= \sum_{k=1}^{n} \mathrm{P}(X = k)\mathrm{P}(Y = k) \qquad \text{(independence)}$$

$$= \sum_{k=1}^{n} p_k q_k.$$

(b) If we knew the value of X, we would be able to find the chance that $Y > X$. In particular, if $X = k$, then the chance that Y is greater than X is $q_{k+1} + \ldots + q_N$.

This is our cue to partition on the value of X. We get that

$$\begin{aligned}
\mathrm{P}(Y > X) &= \sum_{k=1}^{N} \mathrm{P}(Y > X \cap X = k) \\
&= \sum_{k=1}^{N} \mathrm{P}(Y > X \mid X = k)\mathrm{P}(X = k) \\
&= \sum_{k=1}^{N} \mathrm{P}(Y > k)\mathrm{P}(X = k) && \text{(independence)} \\
&= \sum_{k=1}^{N} \left(\sum_{j=k=1}^{N} q_j \right) p_k \\
&= \sum_{k=1}^{N} \sum_{j=k+1}^{N} p_k q_j.
\end{aligned}$$

(c) By the division rule, this is

$$\mathrm{P}(Y > X \mid X = 3) = \frac{\mathrm{P}(Y > X \cap X = 3)}{\mathrm{P}(X = 3)}.$$

This is

$$\frac{\sum_{k=4}^{N} q_k \cdot p_3}{p_3} = \sum_{k=4}^{N} q_k.$$

The same result follows more directly from observing that $\mathrm{P}(Y > X \mid X = 3) = \mathrm{P}(Y > 3 \mid X = 3) = \mathrm{P}(Y > 3)$, where the last equality is from the fact that X and Y are independent.

(d) We have seen in previously and in Question 5 that it is easier to find the complement $\mathrm{P}(\min(X, Y) > n) = \mathrm{P}(X > n, Y > n) = \mathrm{P}(X > n)\mathrm{P}(Y > n)$, where the probability splits into a product because X and Y are independent.

Then

$$\mathrm{P}(\min(X, Y) > n) = \mathrm{P}(X > n)\mathrm{P}(Y > n) = \left(\sum_{k=n+1}^{N} p_k \right) \left(\sum_{k=n+1}^{n} q_k \right).$$

So

$$\mathrm{P}(\min(X, Y) \leq n) = 1 - \mathrm{P}(\min(X, Y) > n) = 1 - \left(\sum_{k=n+1}^{N} p_k \right) \left(\sum_{k=n+1}^{n} q_k \right).$$

Another way is to observe that by inclusion-exclusion,

$$\begin{aligned}
\mathrm{P}(\min(X, Y) \leq n) &= \mathrm{P}(X \leq n \cup Y \leq n) \\
&= \mathrm{P}(X \leq n) + \mathrm{P}(Y \leq n) - \mathrm{P}(X \leq n, Y \leq n) \\
&= \sum_{k=1}^{n} p_k + \sum_{k=1}^{n} q_k - \sum_{k=1}^{n} \sum_{j=1}^{n} p_k q_k.
\end{aligned}$$

(e) By the division rule, this is

$$P(X+Y \leq N \mid X \geq 3) = \frac{P(X+Y \leq N, X \geq 3)}{P(X \geq 3)}.$$

The denominator is

$$P(X \geq 3) = \sum_{k=3}^{N} p_k.$$

To find the numerator, we can partition on the value of X.

$$\begin{aligned}
P(X+Y \leq N, X \geq 3) &= \sum_{k=1}^{N} P(X+Y \leq N, X \geq 3, X = k) \\
&= \sum_{k=1}^{2} P(X+Y \leq N, X \geq 3, X = k) + \sum_{k=3}^{N} P(X+Y \leq N, X \geq 3, X = k) \\
&= \sum_{k=1}^{2} 0 + \sum_{k=3}^{N} P(X+Y \leq N, X = k) \\
&= \sum_{k=3}^{N} P(X+Y \leq N \mid X = k) P(X = k) \\
&= \sum_{k=3}^{N} P(Y + k \leq N) P(X = k) && \text{(independence)} \\
&= \sum_{k=3}^{N} P(Y \leq N - k) P(X = k) \\
&= \sum_{k=3}^{N} \sum_{j=1}^{N-k} P(Y = j) P(X = k) \\
&= \sum_{k=3}^{N} \sum_{j=1}^{N-k} q_j p_k.
\end{aligned}$$

This can also be found by drawing an $N \times N$ grid and considering the possible pairs (x, y) which are in the event $\{X+Y \leq N, X \geq 3\}$. So the answer is

$$P(X+Y \leq N \mid X \geq 3) = \frac{\sum_{k=3}^{n} \sum_{j=1}^{N-k} q_j p_k}{\sum_{k=3}^{N} p_k}$$

5 Minima and Maxima

(a) Suppose $W \leq k$. Then, since $W = \max\{X_1, \ldots, X_n\}$, we have that $X_1 \leq W, \ldots, X_n \leq W$. Therefore, since $W \leq k$, we have that $X_1 \leq k, \ldots, X_n \leq k$, as desired.

Now suppose for contraposition that $W > k$. In particular, suppose that $W = m$ for some $m > k$. Then there must exist some X_i such that $X_i = m > k$. So it is not the case that $X_1 \leq k, \ldots, X_n \leq k$.

(b) By part **(a)**,

$$\begin{aligned}
P(W \leq k) &= P(X_1 \leq k, \ldots, X_n \leq k) \\
&= P(X_1 \leq k) \cdot \ldots \cdot P(X_n \leq k) && \text{(independence)} \\
&= \left(\frac{k}{N}\right) \cdot \ldots \cdot \left(\frac{k}{N}\right) && (X_i \text{ is sampled at random}) \\
&= \left(\frac{k}{N}\right)^n.
\end{aligned}$$

(c) The possible values of W are $\{1, \ldots, N\}$.

For each $k \in \{1, \dots, N\}$, we have that

$$\begin{aligned}
P(W = k) &= P(W \leq k \text{ and } W \not\leq k-1) \\
&= P(\{W \leq k\} \setminus \{W \leq k-1\}) \\
&= P(W \leq k) - P(W \leq k-1) && \text{(difference rule)} \\
&= \left(\frac{k}{N}\right)^n - \left(\frac{k-1}{N}\right)^n
\end{aligned}$$

(d) By a very similar argument to **(a)**,

$$\{V > k\} = \{X_1 > k, \dots, X_n > k\}.$$

Therefore, by the same process, we get that

$$P(V > k) = \left(\frac{N-k}{N}\right)^n.$$

By again applying the difference rule, we get that

$$P(V = k) = P(V > k-1) - P(V > k) = \left(\frac{N-(k-1)}{N}\right)^n - \left(\frac{N-k}{N}\right)^n \qquad k \in \{1, \dots, N\}.$$

(e) We must find $P(V > k)$ for $k \in \{1, \dots N\}$. There are two ways to do this.

For the first way, we observe that there are $N-k$ total elements greater than k, and we want to select n of them. There are $\binom{N}{n}$ total outcomes, of which $\binom{N-k}{n}$ outcomes only have elements greater than $N-k$. So the chance is

$$P(V > k) = \frac{\binom{N-k}{n}}{\binom{N}{n}}.$$

We can also find the chance by considering each draw one at a time. The first draw has $N-k$ options out of N equally likely options, the second draw has $N-k-1$ options out of $N-1$ equally likely options, and so on:

$$P(V > k) = \frac{N-k}{N} \cdot \frac{N-k-1}{N-1} \cdot \ldots \cdot \frac{N-k-(n-1)}{N-(n-1)} = \frac{(N-k)!/(N-k-n)!}{N!/(N-n)!}$$

So we get that

$$\begin{aligned}
P(V = k) &= P(V > k-1) - P(V > k) \\
&= \frac{(N-(k-1))!/(N-(k-1)-n)!}{N!/(N-n)!} - \frac{(N-k)!/(N-k-n)!}{N!/(N-n)!} \\
&= \frac{\binom{N-(k-1)}{n}}{\binom{N}{n}} - \frac{\binom{N-k}{n}}{\binom{N}{n}}.
\end{aligned}$$

6 Class Participation

(a) On each call, the professor calls on a computer science major with probability 0.6, independently of all other calls. We are counting the number of independent and identically distributed trials (calls) until a success (a computer science major). So $X \sim \text{Geometric}(0.6)$.

(b) Let X_3 be the number of calls until the professor calls on a computer science major 3 times.

If it takes more than 10 trials to call on 3 computer science majors, there must have been fewer than 2 computer science majors in the first 10 calls. In particular, let Y_{10} be the number of computer science majors in the first 10 calls. This is counting the number successes (computer science majors) in 10 independent and identically distributed trials (calls). So $Y_{10} \sim \text{Binomial}(10, 0.6)$ and we have that

$$P(X_3 > 10) = P(Y_{10} < 3) = \sum_{k=0}^{2} \binom{10}{k} 0.6^k 0.4^{10-k}.$$

(c) On each call, the professor calls on a computer science or a data science major with chance $0.6 + 0.3 - 0.1 = 0.8$ by inclusion-exclusion. So we are counting the number of independent and identically distributed trials (calls) until a success (a data science major or a computer science major). So $Y \sim \text{Geometric}(0.8)$.

(d) This is $\max(W, X)$ where W is the number of calls it takes to call on a data science major. $W \sim \text{Geometric}(0.3)$ by the reasoning in part **(a)**.

Then, since $X + W = \min(X, W) + \max(X, W)$, we have that $\max(X, W) = X + W - \min(X, W)$ and hence

$$\text{E}[\max(X, W)] = \text{E}[X] + \text{E}[W] - \text{E}[\min(X, W)]$$

by linearity. Note that $Y = \min(X, W)$, so we have that

$$\text{E}[\max(X, W)] = \frac{1}{0.6} + \frac{1}{0.3} - \frac{1}{0.8}.$$

This same expectation can also be found by conditioning on the first call.

7 The Poisson and Binomial Distributions

(a) If there are n trials, X is the number of successes in n independent and identically distributed trials. That's a binomial distribution. So $X \mid N = n \sim \text{Binomial}(n, p)$.

(b) The possible values of X are $\{0, 1, \ldots\}$, since it's possible to get any number of successes as it's possible to get any number of trials.

Given $N = n$, we know the distribution of X. That's our cue to partition on N. Let $q = 1 - p$. We have that, for each $k \in \{0, 1, \ldots\}$,

$$\begin{aligned}
\text{P}(X = k) &= \sum_{n=0}^{\infty} \text{P}(X = k \mid N = n)\text{P}(N = n) \\
&= \sum_{n=0}^{k-1} \text{P}(X = k \mid N = n)\text{P}(N = n) + \sum_{n=k}^{\infty} \text{P}(X = k \mid N = n)\text{P}(N = n) \\
&= \sum_{n=0}^{k-1} 0 \cdot \text{P}(N = n) + \sum_{n=k}^{\infty} \binom{n}{k} p^k q^{n-k} \cdot e^{-\mu} \frac{\mu^n}{n!} \\
&= \sum_{n=k}^{\infty} \frac{n!}{k!(n-k)!} p^k q^{n-k} e^{-\mu} \frac{\mu^n}{n!} \\
&= \frac{e^{-\mu} p^k}{k!} \sum_{n=k}^{\infty} \frac{q^{n-k} \mu^n}{(n-k)!}.
\end{aligned}$$

That infinite sum looks a lot like the Taylor series expansion

$$e^{q\mu} = \sum_{i=0}^{\infty} \frac{(q\mu)^i}{i!}.$$

Re-indexing it yields

$$\sum_{n=k}^{\infty} \frac{q^{n-k}\mu^n}{(n-k)!} = \sum_{i=0}^{\infty} \frac{q^i \mu^{i+k}}{i!} = \mu^k \sum_{i=0}^{\infty} \frac{(q\mu)^i}{i!} = \mu^k e^{q\mu}.$$

Therefore, for $k \in \{0, 1, \ldots\}$,

$$\text{P}(X = k) = \frac{e^{-\mu} p^k}{k!} \cdot \mu^k e^{q\mu} = e^{-\mu + q\mu} \frac{(p\mu)^k}{k!} = e^{-p\mu} \frac{(p\mu)^k}{k!}.$$

That is, $X \sim \text{Poisson}(p\mu)$. Woah!

(c) The conditional distribution of $Y \mid N = n$ is binomial (n, q). By parts **(a)** and **(b)**, this means that $Y \sim \text{Poisson}(q\mu)$.

(d) We check whether $\text{P}(X = k, Y = j) = \text{P}(X = k)\text{P}(Y = j)$. Since we don't understand the dependence between X and Y but do understand the dependence between X and N, we rewrite the event $\{X = k, Y = j\}$ in terms of $\{X = k, N = k + j\}$.

$$\begin{aligned}
\text{P}(X = k, Y = j) &= \text{P}(X = k, N = k + j) \\
&= \text{P}(X = k \mid N = k + j)\text{P}(N = k + j)
\end{aligned}$$

$$\begin{aligned}
&= \binom{k+j}{k} p^k q^j \cdot e^{-\mu} \frac{\mu^{k+j}}{(k+j)!} \\
&= \frac{(k+j)!}{k!j!} p^k q^j e^{-(p+q)\mu} \frac{\mu^k \mu^j}{(k+j)!} \\
&= e^{-p\mu} \frac{(p\mu)^k}{k!} \cdot e^{-q\mu} \frac{(q\mu)^j}{j!} \\
&= \mathrm{P}(X = k)\mathrm{P}(Y = j).
\end{aligned}$$

So X and Y are indeed independent. That's crazy!

8 Billiard Balls

(a) There are two ways to count this. The first way is to observe that this is like throwing two balls into n bins labelled 1 through n. There are $\binom{n-1+2}{2} = \binom{n+1}{2}$ ways to do this.

We can also count by separately considering pairs where $i = j$ and pairs where $i < j$.

(1) There are n pairs where $i = j$: the pairs (i, i) for $i \in \{1, \dots, n\}$.

(2) There are $\binom{n}{2}$ pairs where $i < j$: there are $\binom{n}{2}$ ways to choose a pair of two distinct numbers from $1, \dots, n$, and for each such pair, i will be the smaller of the two numbers and j will be the larger of the two.

So the total number of such ordered pairs is

$$\binom{n+1}{2} = n + \binom{n}{2}.$$

(b) By definition, Charlie chooses his ordered pair uniformly at random from all ordered pairs. So we can find the probability by counting the number of ordered pairs (i, j) such that $i \le k \le j$ and dividing by our answer to part **(a)**.

Any ordered pair (i, j) in which ball k gets painted must satisfy $i \le k$ and $j \ge k$. For i, any of the values $1, \dots, k$ will work, so i has k possible choices. For j, any of the values $k, k+1, \dots, n$ will work, so j has $n - (k-1)$ possible choices.

Therefore, there are $k \cdot (n - (k-1))$ ordered pairs which result in ball k getting painted. By equally likely outcomes, we have that

$$p_k = \mathrm{P}(\text{ball } k \text{ is painted}) = \frac{k \cdot (n - (k-1))}{\binom{n+1}{2}}.$$

(c) There are many ways for ball k to have been painted: it could have been painted on the first round, or the second round, or the third round, and so on. So we'll instead consider the complement: the chance that ball k is never painted in the r rounds. The only way for this to happen is the ball isn't painted in each of the r rounds.

Since each round paints ball k black independently with probability $1 - p_k$, ball k has a $(1 - p_k)^r$ chance of never being painted by the end of round r.

Therefore the chance that ball k is painted at some point in the first r rounds is

$$1 - (1 - p_k)^r.$$

(d) Let $X = I_1 + \dots + I_n$, where I_k is the indicator of the event that ball k is painted after r rounds. Then, by linearity of expectation,

$$\begin{aligned}
\mathrm{E}[X] &= \mathrm{E}[I_1 + \dots + I_n] \\
&= \mathrm{E}[I_1] + \dots + \mathrm{E}[I_n] \\
&= 1 - (1 - p_1)^r + \dots + 1 - (1 - p_n)^r \\
&= n - \sum_{i=1}^{n} (1 - p_i)^r,
\end{aligned}$$

where $\mathrm{E}[I_k] = \mathrm{P}(I_k = 1) = \mathrm{P}(\text{ball } k \text{ painted after } r \text{ rounds}) = 1 - (1 - p_k)^r$.

(e) Finding the distribution of R is quite nasty, nor are we able to decompose R as a sum of indicators or other random variables (we can, in fact, write $R = \max\{X_1, \ldots, X_n\}$ for X_k the time until the k^{th} ball is painted, but that's not too helpful since the X_k are not independent). Nor can we view R as some function of a simpler random variable whose distribution we do know, nor is R a famous distribution that we're familiar with.

All we're left with is the tail-sum formula (which we might have been inclined to use once we observed that R was a maximum).

We have that

$$\mathrm{E}[R] = \sum_{r=0}^{\infty} \mathrm{P}(R > r).$$

To find $\mathrm{P}(R > r)$, observe that if it takes more than r rounds to paint all n balls, there must be at least one ball which was not painted in those r rounds. Let B_k^r be the event that ball k was never painted by the end of r rounds. We have that

$$\mathrm{P}(R > r) = \mathrm{P}\left(\bigcup_{k=1}^{n} B_k^r\right).$$

Note that the events in the union are not mutually exclusive, so we have to use inclusion-exclusion. Let's first consider $\mathrm{P}(B_k^r)$. We could just use part **(b)** to get that

$$\mathrm{P}(B_k^r) = 1 - \mathrm{P}(\text{ball } k \text{ painted after } r \text{ rounds}) = (1 - p_k)^r.$$

However, this will not extend easily to finding $\mathrm{P}(B_k^r \cap B_{k'}^r)$ and later intersections. Instead, observe that if ball k is never painted in r rounds, then in each round we have that Charlie selects $i, j < k$ or $i, j > k$.

(1) Choosing two indices $i, j < k$ is like throwing 2 balls into $k-1$ bins labelled 1 through $k-1$; there are $\binom{(k-1)+1}{2} = \binom{k}{2}$ ways to do this.

(2) Similarly, choosing two indices $i, j > k$ is like throwing two balls into $n - k$ bins labelled $k + 1$ through n; there are $\binom{(n-k)+1}{2}$ ways to do this.

Therefore the chance that ball k is not painted on any given round is

$$\frac{\binom{k}{2} + \binom{n-k+1}{2}}{\binom{n+1}{2}}.$$

Since the rounds are independent, the chance that ball k is not painted after r rounds is

$$\mathrm{P}(B_k^r) = \left(\frac{\binom{k}{2} + \binom{n-k+1}{2}}{\binom{n+1}{2}}\right)^r.$$

To extend this to two balls, we'll need to generalize our notation a little bit. Let $k_1 < k_2$ be two balls. the chance that neither are colored after r rounds is the chance that in each round, Charlie picks $i, j < k_1$, $k_1 < i, j < k_2$, or $i, j > k_2$.

(1) $i, j < k_1$. This is like throwing 2 balls into $k_1 - 1$ bins labelled 1 through $k_1 - 1$. There are $\binom{k_1}{2}$ ways to do this.

(2) $i, j > k_2$. This is like throwing 2 balls into $n - k_2$ balls labelled $k_2 + 1$ through n. There are $\binom{k_2}{2}$ ways to do this.

(3) $k_1 < i, j < k_2$. This is like throwing 2 balls into $k_2 - 1 - k_1$ bins labelled $k_1 + 1$ through $k_2 - 1$. There are $\binom{(k_2-1-k_1)+1}{2} = \binom{k_2-k_1}{2}$ ways to do this.

Therefore on each round, the chance that neither k_1 nor k_2 are painted is

$$\frac{\binom{k_1}{2} + \binom{k_2-k_1}{2} + \binom{n-k_2+1}{2}}{\binom{n+1}{2}}.$$

Since the rounds are independent, the chance that neither is painted after r rounds is

$$\mathrm{P}(B_{k_1}^r \cap B_{k_2}^r) = \left(\frac{\binom{k_1}{2} + \binom{k_2-k_1}{2} + \binom{n-k_2+1}{2}}{\binom{n+1}{2}}\right)^r.$$

Exetnding this, to find the chance that none of $k_1 < \ldots < k_m$ are colored red after r rounds, this same argument gets that this chance is

$$\mathrm{P}\left(\bigcap_{i=1}^{m} B_{k_i}^r\right) = \left(\frac{\binom{k_1}{2} + \sum_{i=2}^{m-1} \binom{k_i-k_{i-1}}{2} + \binom{n+1-k_m}{2}}{\binom{n+1}{2}}\right)^r$$

Therefore, by inclusion-exclusion, we have that

$$\begin{aligned}
\mathrm{P}(R > r) &= \sum_{k_1} \mathrm{P}(B_{k_1}^r) - \sum_{k_1 < k_2} \mathrm{P}(B_{k_1}^r \cap B_{k_2}^r) + \sum_{k_1 < k_2 < k_3} \mathrm{P}(B_{k_1}^r \cap B_{k_2}^r \cap B_{k_3}^r) - \ldots + (-1)^{n-1} \mathrm{P}\left(\bigcap_{i=1}^n B_{k_i}^r\right) \\
&= \sum_{k_1} \left(\frac{\binom{k_1 - 0}{2} + \binom{n+1-k_1}{2}}{\binom{n+1}{2}} \right)^r - \sum_{k_1 < k_2} \left(\frac{\binom{k_1 - 0}{2} + \binom{k_2 - k_1}{2} + \binom{n+1-k_2}{2}}{\binom{n+1}{2}} \right)^r \\
&\quad + \sum_{k_1 < k_2 < k_3} \left(\frac{\binom{k_1 - 0}{2} + \binom{k_2 - k_1}{2} + \binom{k_3 - k_2}{2} + \binom{n+1-k_3}{2}}{\binom{n+1}{2}} \right)^r \\
&\quad - \ldots \\
&\quad + (-1)^{n-2} \sum_{k_1 < \ldots < k_{n-1}} \left(\frac{\binom{k_1 - 0}{2} + \binom{k_2 - k_1}{2} + \ldots + \binom{k_{n-1} - k_{n-2}}{2} + \binom{n+1-k_{n-1}}{2}}{\binom{n+1}{2}} \right)^r \\
&\quad + (-1)^{n-1} \sum_{k_1 < \ldots < k_n} \left(\frac{\binom{k_1 - 0}{2} + \binom{k_2 - k_1}{2} + \ldots + \binom{k_n - k_{n-1}}{2} + \binom{n+1-k_n}{2}}{\binom{n+1}{2}} \right)^r
\end{aligned}$$

This can be written more succintly as

$$\mathrm{P}(R > r) = \sum_{\ell=2}^{n+1} (-1)^{\ell-2} \sum_{\substack{k_0 < \ldots < k_\ell \\ k_0 = 0 \\ k_\ell = n+1}} \left(\sum_{i=1}^{\ell} \frac{\binom{k_i - k_{i-1}}{2}}{\binom{n+1}{2}} \right)^r$$

Therefore, by the tail-sum formula, we have that

$$\mathrm{E}[R] = \sum_{r=0}^{\infty} \mathrm{P}(R > r) = \sum_{r=0}^{\infty} \sum_{\ell=2}^{n+1} (-1)^{\ell-2} \sum_{\substack{k_0 < \ldots < k_\ell \\ k_0 = 0 \\ k_\ell = n+1}} \left(\sum_{i=1}^{\ell} \frac{\binom{k_i - k_{i-1}}{2}}{\binom{n+1}{2}} \right)^r .$$

9 Practicing Expectations

(a) This is the expected number of individuals (varieties) from a population (five total varieties) which meet some criterion (fully in stock at the end of the day). That is our cue to use indicators. Let $I_1, \ldots, I_5$ be indicators such that $I_j = 1$ if the j^{th} variety has all 20 boxes left at the end of the day.

Then $X = I_1 + \ldots + I_5$, so

$$\begin{aligned}
\mathrm{E}[X] &= \mathrm{E}[I_1 + \ldots + I_5] \\
&= \mathrm{E}[I_1] + \ldots + \mathrm{E}[I_5] \\
&= 5\mathrm{E}[I_1],
\end{aligned}$$

where the equality in the last line follows from the fact that $\mathrm{E}[I_1] = \ldots = \mathrm{E}[I_5]$ since the indicators are identically distributed: the varieties are chosen at random, so we cannot have that a specific variety is more or less likely to be chosen by the customers than any other variety.

Then

$$\begin{aligned}
\mathrm{E}[I_1] &= \mathrm{P}(I_1 = 1) \\
&= \mathrm{P}(\text{all 20 boxes of the first variety are left}) \\
&= \frac{\binom{80}{25}}{\binom{100}{25}}.
\end{aligned}$$

The probability here is by the fact that there are $\binom{100}{25}$ equally likely outcomes (consisting of the 25 varieties that were chosen by the customers), of which there are $\binom{80}{25} = \binom{80}{25}\binom{20}{0}$ outcomes in which none of the boxes from the first variety are chosen. Therefore

$$\mathrm{E}[X] = 5 \frac{\binom{80}{25}}{\binom{100}{25}}.$$

(b) We know the distribution of X and are trying to find the expectation of $|X-5|$, which is a function of X. That's our cue to use the law of the unconscious statistician (also known as the nonlinear function rule). We have that

$$\begin{aligned}
\mathrm{E}[|X-5|] &= \sum_{i=1}^{9} |i-5|\mathrm{P}(X=i) \\
&= \frac{1}{9}\sum_{i=1}^{9} |i-5| \\
&= \frac{1}{9}\Big(|1-5|+|2-5|+|3-5|+|4-5|+|5-5|+|6-5|+|7-5|+|8-5|+|9-5|\Big) \\
&= \frac{1}{9}\Big(4+3+2+1+0+1+2+3+4\Big) \\
&= \frac{2(10)}{9}.
\end{aligned}$$

(c) As we have seen in Question 5, we can easily find the tails of minima and maxima. So we will use the tail sum formula. Let $X_1, \ldots, X_n$ be the n values drawn and $V = \min\{X_1, \ldots, X_n\}$ be their minimum. We have by Question 5**(d)** that, for $k \in \{1, \ldots, N\}$,

$$\mathrm{P}(V > k) = \left(\frac{N-k}{N}\right)^n,$$

so the expectation is

$$\mathrm{E}[V] = \sum_{k=0}^{\infty} \mathrm{P}(V>k) = \sum_{k=0}^{N}\left(\frac{N-k}{N}\right)^n + \sum_{k=N+1}^{\infty} 0 = \sum_{k=0}^{N}\left(\frac{N-k}{N}\right)^n.$$

Alternatively, we can just use the definition of expectation in conjunction with the distribution for V which we found in Question 5**(d)**.

$$\mathrm{E}[V] = \sum_{k=1}^{N} k\mathrm{P}(V=k) = \sum_{k=1}^{n} k\left(\left(\frac{N-(k-1)}{N}\right)^n - \left(\frac{N-k}{N}\right)^n\right)$$

(d) We are finding the expected number of individuals (faces) in a population (six faces of a die) which meet a certain criterion (appear exactly twice), so that's our cue to use indicators. Let $I_1, \ldots, I_6$ be the indicators such that $I_j = 1$ if the j^{th} face appears exactly twice. Then for X the random variable representing Casey's score, $X = I_1 + \ldots + I_6$, and so

$$\mathrm{E}[X] = \mathrm{E}[I_1 + \ldots + I_6] = \mathrm{E}[I_1] + \ldots + \mathrm{E}[I_6] = 6\mathrm{E}[I_1],$$

where the last equality follows from the fact that $I_1, \ldots, I_6$ are identically distributed since no face is more or less likely to appear twice than any other face.

Then

$$\mathrm{P}(I_1 = 1) = \mathrm{P}(\text{first face appears exactly twice}).$$

Let N_1 be the number of times the first face appears in n rolls. This is the number of successes (rolling the first face) in independent and identically distributed trials (n rolls), so $N_1 \sim \text{Binomial}(n, 1/6)$. Therefore

$$\mathrm{P}(I_1 = 1) = \mathrm{P}(N_1 = 2) = \binom{n}{2}\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^{n-2}.$$

So we have that

$$\mathrm{E}[X] = 6\mathrm{P}(I_1 = 1) = 6\binom{n}{2}\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^{n-2}.$$

(e) If we knew the value of X, we would be able to find the expected number of heads. That's our cue to use conditioning.

For starters, note that if $p = 0$, then Xiayi sees 0 heads. So consider the case where $p > 0$.

Let N be the total number of heads which Xiayi sees. Then $N = 1 + Y$, where Y is the number of heads which Xiayi sees in the X tosses of the coin. We have that $\mathrm{E}[N] = 1 + \mathrm{E}[Y]$.

To find $\mathrm{E}[Y]$, we can condition on the value of X. Given $X = x$, Y is a binomial (k, p) distribution. So $\mathrm{E}[Y \mid X = x] = xp$. Note that this is a function of X: we have that $\mathrm{E}[Y \mid X = x] = g(x) = xp$. Therefore $\mathrm{E}[Y \mid X] = g(X) = Xp$.

Then by iterated expectation,

$$\mathrm{E}[Y] = \mathrm{E}[\mathrm{E}[Y \mid X]] = \mathrm{E}[Xp] = \mathrm{E}[X]p = \frac{1}{p} \cdot p = 1.$$

Therefore the expected number of heads Xiayi sees is $1 + 1 = 2$. Our answer is

$$\mathrm{E}[N] = \begin{cases} 0 & p = 0, \\ 2 & \text{otherwise.} \end{cases}$$