Problem 1

Suppose an urn has \( b \) blue balls and \( r \) red balls. We randomly pick a ball. If the ball is red, we put two red balls back to the urn. If the ball is blue, we put two blue balls back into the urn. Then we randomly pick a ball again.

The problem is a bit awkwardly worded. We draw a ball. If the ball we draw is red, then we throw out the ball we chose, and two red balls fall out of the sky and into the urn. The same for the blue balls. With that said, on each draw, the number of balls in the urn grows.

(1) **What is the probability that the first pick is red?**

The first pick is only restricted to the balls that are in the urn. There are \( b \) blue balls, \( r \) red balls and \( b + r \) total balls. Thus,

\[
P(\text{first ball is red}) = \frac{r}{r + b}
\]

(2) **What is the probability that the second pick is red?**

To get to the second pick being red, we must have a first pick. There are two ways we can get the second pick being red. Either the first pick is red, or the first pick is blue. These events are mutually exclusive. Let \( A \) be the event that the second pick is red. Let \( B \) be the event that the first pick is red. Then, \( A' \) is the event that the second pick is blue and \( B' \) is the event that the first pick is blue.

The problem can be formulated as

\[
P(\text{second pick is red}) = P(\text{second pick is red} \cap \text{first pick is red}) + P(\text{second pick is red} \cap \text{first pick is blue})
\]

\[
P(A) = P(A \cap B) + P(A \cap B')
\]

which is known as the **law of total probability** and is given by:

\[
P(A) = \sum_k P(A \cap B_k) = \sum_k P(A|B_k)P(B_k)
\]

Since the number of balls in the urn on the second pick is dependent on the first pick, the events are not independent. We must use the law of total probability.

\[
P(A) = P(A \cap B) + P(A \cap B')
\]

\[
= P(A|B)P(B) + P(A|B')P(B')
\]

- \( P(A|B) \) is the probability that the second pick is red given that the first pick is red. We know that the first pick is red. This means, there is one less red ball (the one we chose), and we add in two more red balls into the urn for the second draw and leave the number of blue balls the same. The number of red balls is now \( r + 1 \) and the total number of balls is now \( r + b + 1 \). Thus \( P(A|B) = \frac{r+1}{r+b+1} \).
• $P(B)$ is the probability that the first pick is red. This comes from the first part of the problem. $P(A) = \frac{r}{r + b}$.

• $P(A|B')$ is the probability that the second pick is red given that the first pick is blue. We know that the first pick is blue. This means, there is one less blue ball (the one we chose), and we add in two more blue balls into the urn for the second draw and leave the number of red balls the same. The number of blue balls is now $b + 1$ and the total number of balls is now $r + b + 1$. Thus, $P(A|B') = \frac{r}{r + b + 1}$.

• $P(B')$ is the probability that the first pick is blue. Trivially, $P(B') = 1 - P(B) = \frac{b}{r + b}$.

Then we have that

\[
P(A) = \frac{r + 1}{r + b + 1} \cdot \frac{r}{r + b} + \frac{r}{r + b + 1} \cdot \frac{b}{r + b}
\]
\[
= \frac{r + 1}{r + b + 1} \cdot \frac{r}{r + b} + \frac{r}{r + b + 1} \cdot \frac{b}{r + b}
\]
\[
= \frac{r(r + b + 1)}{(r + b + 1)(r + b)}
\]
\[
= \frac{r}{r + b + 1}
\]
\[
= P(B)
\]

Interesting! We see that the probability of the second pick being red is equal to the probability of the first pick being red!!! Hmm...

**Aside:** In this problem, the number of balls in the urn increases on each draw. Had we not added balls into the urn, we would have *sampling without replacement* and the problem would resemble the hypergeometric distribution.

(3) *(Optional)* If we continue, then what is the probability that the third pick is red?

This problem is known as the *Polya Urn Scheme* and it has many uses, one of which is the modeling of infectious diseases. It also has use in graph theory and network analysis in the theory of preferential attachment. Because of the previous use, it has also been used to study evolutionary processes. Another use is in signal processing and image processing. A matter of fact, there is an entire book about this type of problem!

To answer this part, we have to do many tedious calculations using the *chain rule*. First, use the law of total probability on the third draw being red. We need to consider every possible case where the third draw is red.

\[
P(\text{third is red}) = P(\text{third is red} \cap \text{first is red} \cap \text{second is red})
\]
\[
= + P(\text{third is red} \cap \text{first is red} \cap \text{second is blue})
\]
\[
= + P(\text{third is red} \cap \text{first is blue} \cap \text{second is red})
\]
\[
= + P(\text{third is red} \cap \text{first is blue} \cap \text{second is blue})
\]
Now we need to convert these to conditionals. For two draws, we would use the “Bayes trick.” This is actually just a trivial case of the **chain rule**! The third draw needs to take into account the first and second draws. Thus, by the chain rule, we get

\[
P(\text{third is red}) = P(\text{third is red} | \text{second is red } \cap \text{ first is red})P(\text{second is red} | \text{first is red})P(\text{first is red})
\]
\[
+ P(\text{third is red} | \text{second is red } \cap \text{ first is blue})P(\text{second is red} | \text{first is blue})P(\text{first is blue})
\]
\[
+ P(\text{third is red} | \text{second is blue } \cap \text{ first is red})P(\text{second is blue} | \text{first is red})P(\text{first is red})
\]
\[
+ P(\text{third is red} | \text{second is blue } \cap \text{ first is blue})P(\text{second is blue} | \text{first is blue})P(\text{first is blue})
\]

Note that all four terms in the above sum each contain three parts. The second two parts \(P(\text{third is red} | \text{first is } \_ )\) and \(P(\text{first is } \_ )\) come from the previous parts of the problem:

\[
P(\text{second is red} | \text{first is red}) = \frac{r + 1}{r + b + 1}
\]
\[
P(\text{second is blue} | \text{first is red}) = \frac{b}{r + b + 1}
\]
\[
P(\text{second is red} | \text{first is blue}) = \frac{r}{r + b + 1}
\]
\[
P(\text{second is blue} | \text{first is blue}) = \frac{b + 1}{r + b + 1}
\]

We also know from a previous part that \(P(\text{first is red}) = \frac{r}{r + b}\) and \(P(\text{first is blue}) = 1 - \frac{r}{r + b} = \frac{b}{r + b}\).

The monster terms are the only ones that require thinking – the terms that look like \(P(\text{third is red} | \text{second is } \_ \cap \text{ first is } \_ )\). We need to think about the state the urn is in after the first two draws.

- To compute \(P(\text{third is red} | \text{second is red } \cap \text{ first is red})\), this means we must have drawn a red on the first draw. This means we throw out that ball and put two red balls back into the urn. The number of red balls is then \(r - 1 + 2 = r + 1\). The total number of balls is then \(r + b - 1 + 2 = r + b + 1\). If the second draw is also red, then we throw out the ball, and put two red balls back into the urn. Now the total number of red balls is \((r + 1) - 1 + 2 = r + 2\) and the total number of balls in the urn is \((r + b + 1) - 1 + 2 = r + b + 2\). Thus, \(P(\text{third is red} | \text{second is red } \cap \text{ first is red}) = \frac{r + 2}{r + b + 2}\).

- To compute \(P(\text{third is red} | \text{second is red } \cap \text{ first is blue})\), this means we must have drawn a blue on the first draw. This means we throw out that ball and put two blue balls back into the urn. The number of blue balls is then \(b - 1 + 2 = b + 1\). The total number of balls is then \(r + b - 1 + 2 = r + b + 1\). If the second draw is red, then we throw out the ball, and put two red balls back into the urn. Now the total number of red balls is \(r - 1 + 2 = r + 1\) and the total number of balls in the urn is \((r + b + 1) - 1 + 2 = r + b + 2\). Thus, \(P(\text{third is red} | \text{second is red } \cap \text{ first is blue}) = \frac{r + 1}{r + b + 2}\).

- To compute \(P(\text{third is red} | \text{second is blue } \cap \text{ first is blue})\), this means we must have drawn a blue on the first draw. This means we throw out that ball and put two blue balls back into the urn. The number of blue balls is then \(b - 1 + 2 = b + 1\). The total number of balls is then \((r + b) - 1 + 2 = r + b + 2\). If the second draw is also blue, then we throw out the ball, and put two blue balls back into the urn. Now the total number of blue balls is
\[(b+1) - 1 + 2 = b+2\] and the total number of balls in the urn is \[(r+b+1) - 1 + 2 = r+b+2.\]

Thus, \[P(\text{third is red} | \text{second is blue} \cap \text{first is blue}) = \frac{r + 1}{r + b + 2}.\]

- To compute \[P(\text{third is red} | \text{second is blue} \cap \text{first is red}),\] this means we must have drawn a red on the first draw. This means we throw out that ball and put two red balls back into the urn. The number of red balls is then \[r - 1 + 2 = r + 1.\]

The total number of balls is then \[r + b - 1 + 2 = r + b + 1.\]

If the second draw is blue, then we throw out the ball, and put two blue balls back into the urn. Now the total number of blue balls is \[b - 1 + 2 = b + 1\] and the total number of balls in the urn is \[(r + b + 1) - 1 + 2 = r + b + 2.\]

Thus, \[P(\text{third is red} | \text{second is blue} \cap \text{first is red}) = \frac{r + 1}{r + b + 2}.\]

Then we get:

\[
P(\text{third is red}) = \frac{r + 2}{r + b + 2} \cdot \frac{r + 1}{r + b + 1} \cdot \frac{r}{r + b} + \frac{r + 1}{r + b + 2} \cdot \frac{r}{r + b + 1} \cdot \frac{b}{r + b}
\]

\[
+ \frac{r + 1}{r + b + 2} \cdot \frac{b}{r + b + 1} \cdot \frac{r}{r + b} + \frac{r}{r + b} \cdot \frac{r}{r + b + 1} \cdot \frac{b}{r + b}
\]

\[
= \frac{(r + 2)(r + 1)r + rb(r + 1) + rb(r + 1) + rb(b + 1)}{(r + b + 2)(r + b + 1)(r + b)}
\]

By Mathematica...

\[
= \frac{r}{r + b}
\]

\[
= P(\text{second is red})
\]

\[
= P(\text{first is red})
\]

Aside: It can be proved by induction that no matter what pick we consider, the probability that the \(n\)th draw will be red is \(\frac{r}{r + b}\). This is a beautiful result.

Unlike a Markov process where the value at time \(t\) depends only on the time at \(t - 1\), the Polya urn scheme yields the same result for all \(t\), even despite its messy calculation. This type of process is called a martingale.

For the curious (and those taking Stats 102A), here is a simulation that constructs an urn of 10 balls, and computes the probability that the third ball drawn is red. The input is a number that denotes the number of balls in the urn that are red. The straight line of the plot shows that regardless of what numbers of red balls we start with, the probability of drawing a red ball on the third draw is the same!

```r
balls <- 10 #number of balls in urn initially
cycles <- 100 #Number of iterations to run.
N <- 1000

result <- matrix(rep(0, 11*3),ncol=3)
for (r in seq(0,10)) {
  distribution <- matrix(rep(0, cycles*3),ncol=3)
  for (cycle in 1:cycles) {
    #Simulate three draws a large number of times, say 100,000 times.
    times.draw.3.is.red <- times.draw.1.is.red <- 0
    for (i in 1:N) { #We are going to simulate drawing 3 balls, N times.
      urn <- c(rep("R", r), rep("B", balls-r))
      drawn.balls <- rep("", 3)
      for (draw in 1:3) {
        drawn.balls[draw] <- sample(urn, size=1)
        urn <- c(urn, drawn.balls[draw])
      }
      times.draw.1.is.red <- times.draw.1.is.red + as.numeric(drawn.balls[1] == "R")
      times.draw.3.is.red <- times.draw.3.is.red + as.numeric(drawn.balls[2] == "R")
    }
    distribution[r, 1] <- times.draw.3.is.red/N
    distribution[r, 2] <- times.draw.1.is.red/N
    distribution[r, 3] <- balls/N
  }
  result[r,] <- distribution
}
```

```r

= \frac{r}{r + b}

= P(\text{second is red})

= P(\text{first is red})

```
if (drawn.balls[3] == "R") {
    times.draw.3.is.red <- times.draw.3.is.red + 1
}
if (drawn.balls[1] == "R") {
    times.draw.1.is.red <- times.draw.1.is.red + 1
}
distribution[cycle,] <- c(r, times.draw.1.is.red/N, times.draw.3.is.red/N)
result[r+1,] <- c(r, mean(distribution[,2]), mean(distribution[,3]))

Problem 2 Suppose a person performs a random walk over three states 1, 2, 3. He starts from state 1. Each step, regardless of the past history, he stays where he is with probability 1/2, and he moves to each one of the other two states with probability 1/4. Let $X_t$ be the state of this person at time $t$, and $X_0 = 1$.

The best way to solve this problem and its subparts is to draw a diagram.

Note that the state at time $t+1$ depends only on the state at time $t$. This is called a Markov process, and the sequence of states a Markov chain. Markov chains are very important for processes in Computer Science (see CS 112) and Statistics (see Stats 102C/202C). More about this later.
(1) What is the distribution of $X_1$, i.e. what is the probability that $X_1 = i$ for $i = 1, 2, 3$?

The random walker starts at position 1, that is $X_0 = 1$. The probability that he will be in state 1 on the next step (doesn’t move, $X_1 = X_0$) is $\frac{1}{2}$. The probability that he will be in state 2 is $\frac{1}{4}$ and state 3 is $\frac{1}{4}$. This comes directly from the problem specification; the probability that he moves to state 2 or state 3 is $\frac{1}{4}$ each (moves to new state, $X_1 \neq X_0$). Thus,

<table>
<thead>
<tr>
<th>$k$</th>
<th>$P(X_1 = k)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$\frac{1}{2}$</td>
</tr>
<tr>
<td>2</td>
<td>$\frac{1}{4}$</td>
</tr>
<tr>
<td>3</td>
<td>$\frac{1}{4}$</td>
</tr>
</tbody>
</table>

(2) What is the distribution of $X_2$?

$X_2$ is dependent on $X_1$, so we need to use the possible states of $X_1$. Suppose we compute $P(X_2 = 1)$. There are three ways that we can get $X_2 = 1$: $X_1 = 1, X_1 = 2$ or $X_1 = 3$. Thus we would compute

$$P(X_2 = 1) = P(X_2 = 1 \cap X_1 = 1) + P(X_2 = 1 \cap X_1 = 2) + P(X_2 = 1 \cap X_1 = 3)$$

This is called the law of total probability and can be expressed as follows

$$P(A) = \sum_k P(A \cap B_k) = \sum_k P(A|B_k)P(B_k)$$

where $A : X_2 = i$ and $B : X_1 = j$. The righthand side is what I have been referring to as the “Bayes trick.” Thus, in general,

$$P(X_2 = k) = \sum_{j \in \{1,2,3\}} P(X_2 = k \cap X_1 = j) = \sum_{j \in \{1,2,3\}} P(X_2 = k|X_1 = j)P(X_1 = j)$$

Thus the computation proceeds as follows

$$P(X_2 = 1) = P(X_2 = 1 \cap X_1 = 1) + P(X_2 = 1 \cap X_1 = 2) + P(X_2 = 1 \cap X_1 = 3)$$

$$= P(X_2 = 1|X_1 = 1)P(X_1 = 1) + P(X_2 = 1|X_1 = 2)P(X_1 = 2) + P(X_2 = 1|X_1 = 3)P(X_1 = 3)$$

$$= \frac{1}{2} \cdot \frac{1}{2} + \frac{1}{4} \cdot \frac{1}{4} + \frac{1}{4} \cdot \frac{1}{4}$$

$$= \frac{1}{4} + \frac{1}{16} + \frac{1}{16}$$

$$= \frac{6}{16} = \frac{3}{8}$$
\[
P(X_2 = 2) = P(X_2 = 2 \cap X_1 = 1) + P(X_2 = 2 \cap X_1 = 2) + P(X_2 = 2 \cap X_1 = 3)
= P(X_2 = 2|X_1 = 1)P(X_1 = 1) + P(X_2 = 2|X_1 = 2)P(X_1 = 2)
\quad + P(X_2 = 2|X_1 = 3)P(X_1 = 3)
= \frac{1}{4} \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{4} + \frac{1}{4} \cdot \frac{1}{4}
= \frac{1}{8} + \frac{1}{8} + \frac{1}{16}
= \frac{5}{16}
\]

\[
P(X_2 = 3) = P(X_2 = 3 \cap X_1 = 1) + P(X_2 = 3 \cap X_1 = 2) + P(X_2 = 3 \cap X_1 = 3)
= P(X_2 = 3|X_1 = 1)P(X_1 = 1) + P(X_2 = 3|X_1 = 2)P(X_1 = 2)
\quad + P(X_2 = 3|X_1 = 3)P(X_1 = 3)
= \frac{1}{4} \cdot \frac{1}{2} + \frac{1}{4} \cdot \frac{1}{4} + \frac{1}{2} \cdot \frac{1}{4}
= \frac{1}{8} + \frac{1}{16} + \frac{1}{8}
= \frac{5}{16}
\]

Thus,

<table>
<thead>
<tr>
<th>(k)</th>
<th>(P(X_2 = k))</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(\frac{5}{16})</td>
</tr>
<tr>
<td>2</td>
<td>(\frac{1}{16})</td>
</tr>
<tr>
<td>3</td>
<td>(\frac{5}{16})</td>
</tr>
</tbody>
</table>
(3) *(Optional) What is the distribution of $X_3$?*

This problem is similar to the other optional problem, but messier in mechanics. Start with the case $X_3 = 1$. We need to enumerate all of the cases where $X_3 = 1$ and use the law of total probability.

$$P(X_3 = 1) = P(X_3 = 1 \cap X_1 = 1 \cap X_2 = 1) + P(X_3 = 1 \cap X_1 = 1 \cap X_2 = 2) + P(X_3 = 1 \cap X_1 = 1 \cap X_2 = 3)$$

Then by using the chain rule

$$P(X_3 = 1) = P(X_3 = 1|X_1 = 1 \cap X_2 = 1)P(X_2 = 1|X_1 = 1)P(X_1 = 1)$$

But this is a Markov process. That is, the state at time $t$ depends only on the state at time $t - 1$, so we can eliminate the $X_1$ term in the assumptions.

$$P(X_3 = 1) = P(X_3 = 1|X_2 = 1)P(X_2 = 1|X_1 = 1)P(X_1 = 1)$$

And by using the result from the previous parts, the result becomes trivial.

$$P(X_3 = 1) = \frac{1}{2} \left( \frac{1}{2} \right) + \frac{1}{4} \left( \frac{1}{4} \right) + \frac{1}{4} \left( \frac{1}{2} \right) + \frac{1}{4} \left( \frac{1}{2} \right) + \frac{1}{2} \left( \frac{1}{4} \right)$$

$$= \frac{1}{8} + \frac{6}{32} + \frac{2}{64} = \frac{22}{64}$$
I could continue and show the other two cases $P(X_3 = 2)$ and $P(X_3 = 3)$ but I think you get the point...or I hope you get the point. We get the following distribution for $X_3$

<table>
<thead>
<tr>
<th>$k$</th>
<th>$P(X_3 = k)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$\frac{22}{64}$</td>
</tr>
<tr>
<td>2</td>
<td>$\frac{21}{64}$</td>
</tr>
<tr>
<td>3</td>
<td>$\frac{21}{64}$</td>
</tr>
</tbody>
</table>

(4) **What is** $P(X_2 = j|X_1 = i)$ and $P(X_1 = i|X_2 = j)$ for $i, j = 1, 2, 3$.

Do not overthink this problem. There are a couple of ways to solve this problem: understanding what is being asked, and brute-force by formula. They both eventually yield the same answer, but one is more useful. Note that the probability that we stay in the same position from time $t$ to time $t + 1$ is $\frac{1}{2}$. The probability of hopping to one of the other states is $\frac{1}{4}$. Thus, it is trivial that

$$P(X_2 = j|X_1 = i) = \begin{cases} \frac{1}{2}, & \text{for } i = j \\ \frac{1}{4}, & i \neq j \end{cases}$$

The other way is to use Bayes rule. First, note that for $i, j = 1, 2, 3$

$$P(X_2 = j|X_1 = i) = \frac{P(X_2 = j \cap X_1 = i)}{P(X_1 = i)} = \frac{P(X_1 = i)P(X_2 = j)}{P(X_1 = i)} = \frac{P(X_2 = j|X_1 = i)P(X_1 = i)}{P(X_1 = i)}$$

But wait a minute! We do not know what $P(X_1 = i|X_2 = j)$ is! Besides, that is the second part of the problem...

Now we must find $P(X_1 = i|X_2 = j)$. Again, we can do this brute-force, or by noticing something clever. By Bayes rule,

$$P(X_1 = i|X_2 = j) = \frac{P(X_1 = i \cap X_2 = j)}{P(X_2 = j)} = \frac{P(X_2 = j|X_1 = i)P(X_1 = i)}{P(X_2 = j)}$$

Applying the result from the first part of this problem, we get

$$P(X_1 = i|X_2 = j) = \begin{cases} \frac{P(X_1 = i)}{2P(X_2 = j)} , & \text{if } i = j \\ \frac{P(X_1 = i)}{4P(X_2 = j)} , & \text{if } i \neq j \end{cases}$$

Then compute $P(X_1 = i|X_2 = j)$ for $i, j = 1, 2, 3$ by using the result from the past two parts. For conciseness, I skip all of the computation as it is just a result of plugging in values from the previous parts of the problem.

| $i$ | $j$ | $P(X_1 = i|X_2 = j)$ |
|-----|-----|---------------------|
| 1   | 1   | $\frac{1}{4}$      |
| 1   | 2   | $\frac{1}{4}$      |
| 1   | 3   | $\frac{1}{4}$      |
| 2   | 1   | $\frac{1}{4}$      |
| 2   | 2   | $\frac{1}{4}$      |
| 2   | 3   | $\frac{1}{4}$      |
| 3   | 1   | $\frac{1}{4}$      |
| 3   | 2   | $\frac{1}{4}$      |
| 3   | 3   | $\frac{1}{4}$      |
\[
\sum_i P(X_2 = i | X_1 = j) = 1
\]

(5) **Aside:** What is the distribution of \(X_N\), where \(N\) is a large number?

We can find the so-called stationary, or steady-state probability of being in any one of the three states. If we make some assumptions about the underlying Markov chain, we can find the stationary probabilities, \(P(X_N = i)\). We define a transition matrix which is simply the probability of moving from state \(i\) to state \(j\) in exactly one move. Elements \(A_{ji}\) represent the probability of moving from state \(i\) to state \(j\). For this problem,

\[
A = \begin{pmatrix}
\frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\
\frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\
\frac{1}{4} & \frac{1}{2} & \frac{1}{4}
\end{pmatrix}
\]

We define a vector \(v\) for the initial state \(X_0\) and put a 1 as the first element, because \(X_0 = 1\) with probability 1. Then, \(P(X_n = i)\) can be computed for any \(n\), including \(n = 1, 2, 3\) as in the previous problems as

\[
P(X_n = i) = (Av)^n = A^n v
\]

by linearity (see Math 33A/115A).

For \(n = 1\):

\[
P(X_1 = i) = Av = \begin{pmatrix}
\frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\
\frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\
\frac{1}{4} & \frac{1}{2} & \frac{1}{4}
\end{pmatrix} \begin{pmatrix}
1 \\
0 \\
0
\end{pmatrix} = \begin{pmatrix}
\frac{1}{2} \\
\frac{1}{4} \\
\frac{1}{4}
\end{pmatrix}
\]

For \(n = 2\):

\[
P(X_2 = i) = A^2 v = \begin{pmatrix}
\frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\
\frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\
\frac{1}{4} & \frac{1}{2} & \frac{1}{4}
\end{pmatrix} \begin{pmatrix}
\frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\
\frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\
\frac{1}{4} & \frac{1}{2} & \frac{1}{4}
\end{pmatrix} \begin{pmatrix}
1 \\
0 \\
0
\end{pmatrix} = \begin{pmatrix}
\frac{6}{16} \\
\frac{6}{16} \\
\frac{6}{16}
\end{pmatrix}
\]

And for \(n = 3\):

\[
P(X_3 = i) = A^3 v = \begin{pmatrix}
\frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\
\frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\
\frac{1}{4} & \frac{1}{2} & \frac{1}{4}
\end{pmatrix} \begin{pmatrix}
\frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\
\frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\
\frac{1}{4} & \frac{1}{2} & \frac{1}{4}
\end{pmatrix} \begin{pmatrix}
\frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\
\frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\
\frac{1}{4} & \frac{1}{2} & \frac{1}{4}
\end{pmatrix} \begin{pmatrix}
1 \\
0 \\
0
\end{pmatrix} = \begin{pmatrix}
\frac{22}{64} \\
\frac{22}{64} \\
\frac{22}{64}
\end{pmatrix}
\]

For \(n = N\), where \(N\) is some very large number:

\[
P(X_N = i) = A^n v = \begin{pmatrix}
\frac{1}{3} \\
\frac{1}{3} \\
\frac{1}{3}
\end{pmatrix}
\]
Google uses this principle to compute the PageRank score of every webpage on the Internet. A large $N \times N$ matrix contains the probability that a person visits some page $j$ after visiting page $i$, where $N$ is the number of web pages on the Internet. Trivially, this matrix is very sparse (it contains mostly zeroes). The so called “random surfer” assumption assumes that a user indiscriminantly chooses a link on the page with an equal probability. The PageRank essentially gives us the probability that a random surfer ends up at some page $p$. By the assumptions of the Markov chain (Statistics 102C/202C), it does not even matter what the initial state is! The PageRank is also defined as the principal eigenvector of the transition matrix. And you thought linear algebra had no use? PageRank is a measure of authority of a web page and is named after its developer, Larry Page, not the fact that it ranks web pages.

This description is solely for fun. On your assignment and the exam, you should stick to the basic probability axioms!

**Problem 3**

*Suppose during a class, the probability that there is a fire in the classroom is $\alpha$. If there is a fire, the probability we hear the fire alarm is $\beta$. If there is not a fire, the probability that there is a fire alarm is $\gamma$. Given that we hear the fire alarm, what is the probability that there is fire?*

We have two events here. Let $A$ be the event that the classroom is on fire. Let $B$ be the event that we hear the fire alarm. Note that the problem is worded awkwardly. We are to assume that if the fire alarm sounds, we hear it. Thus we know the following information

$$P(A) = \alpha, P(B|A) = \beta, P(B|A') = \gamma$$

Note that $P(B|A')$ is also called a *false positive* – the alarm suggests that there is a fire, but there is not a fire. Conversely, a *false negative* would be $P(B'|A)$, the event that the fire alarm doesn’t sound and there is really a fire. Trivially, a false negative would be much worse.

We want to find

$$P(\text{Fire}|\text{Hear Alarm}) = P(A|B)$$

By Bayes rule we have

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Now what do we do? $A$ and $B$ are not independent, by design. If the events were independent, then a fire alarm would have no purpose. Lets start with the numerator. Since $A$ and $B$ are not independent,

$$P(A \cap B) \neq P(A)P(B)$$

So we need to use the Bayes rule trick. Recall from class that

$$P(A \cap B) = P(A|B)P(B) = P(B|A)P(A)$$
Then the numerator becomes

$$P(A \cap B) = P(B|A)P(A) = \beta \alpha$$

(We don’t use the other option because that is what we are trying to compute!) Now consider the denominator $P(B)$ which is the probability that we hear a fire alarm. Note that we have no idea what this probability is a priori, but we can compute it. **We must consider the possible ways that we hear the fire alarm considering the only other information we know is whether or not there is a fire.**

There are two situations when we hear the fire alarm: when there is a fire, and when there is not a fire. Thus,

$$P(B) = P(B \cap A) + P(B \cap A')$$

This is called the **law of total probability** and can be expressed as follows

$$P(A) = \sum_k P(A \cap B_k) = \sum_k P(A|B_k)P(B_k)$$

Using the law of total probability (and the Bayes trick) we get that

$$P(B) = P(B \cap A) + P(B \cap A')$$

$$= P(B|A)P(A) + P(B|A')P(A')$$

$$= \beta \alpha + \gamma (1 - \alpha)$$

Then we get our answer,

$$P(A|B) = \frac{\alpha \beta}{\alpha \beta + \gamma (1 - \alpha)} = \frac{\alpha \beta}{\gamma + \alpha (\beta - \gamma)}$$