Preface:

I took my first probability and stochastic processes course (UC Berkeley's EE 126) in the semester of Spring 2015. In an attempt to understand the material more thoroughly, I began typing up this manuscript of explained concepts and relevant problems and examples, in order to help other students who will take EE 126 and its equivalent course across all universities, as well as to remedy all gaps that may remain in my own knowledge. This booklet is by no means complete.

Chapter 4 consists of homework and recitation problems from the EE 126 course. Some problems have been extended beyond what was taught in the course. The other chapters consist of various concepts that have been explained in several notes from UC Berkeley, MIT, Harvard, and online open courseware from Youtube. A lot of the material in this part of the manuscript are outside of the scope of EE 126 and are more within scope of the sequel graduate course to EE 126 (which is EE 226), especially Martingales, Brownian Motion, Wide Sense Stationary Processes, Gaussian Random Processes, and White Noise. It is my belief that students who plan to take this course as well will find the material of great use. Some examples of practical usage are open ended questions, so they are left partly blank. Nevertheless, an attempted explanation of the solutions have been provided.

The following texts have been consulted in the making of this manuscript so far:

1. Bertsekas, Dimitri, and John N. Tsitsiklis. Introduction to Probability. N.p.: Athena Scientific, n.d. Print.

2. Ross, Sheldon. A First Course in Probability. N.p.: Prentice Hall, n.d. Print.

3. Brzezniak, Zdzislaw, and Tomasz Zastawniak. Basic Stochastic Processes. N.p.: Springer, n.d. Print.

4. Ross, Sheldon. Introduction to Probability Models. N.p.: Academic Press, n.d. Print.

Contents:

Chapter 1: Introduction

Let us begin by first considering a few well-known interesting problems, the Monty Hall problem, the Gambler's Ruin problem, and the Coupon Collecting problem, to give some reason why probability and stochastic processes are useful things to study.

Monty Hall Problem

The formulation is as follows. In a game show, there are three doors labeled 1, 2, and 3. Behind one door is the desired prize, a brand new car. Behind each of the other doors, there is a goat. The player must correctly guess which door the prize is behind in order to win the game (and get the new car).

Suppose the player chooses a door. Afterwards, the game show host opens one of the other two doors to reveal a goat. He then gives the player a chance to reevaluate his decision. Given these events, is it better for the player to stick to his original door, or move to the last unopened door? Namely, what is the probability of winning under each strategy?

First, we will build a sample space $\Omega_i, i = 1, 2, 3$ for each door i that the player initially chooses.

Compute the probability of obtaining the prize for the strategy where the player switches to the other door:

* Suppose the car was behind door 1 and the player chooses door 1. Suppose the host opens door 2. Then under the switching strategy, there is no way that the player will win. The same case applies if the host opens door 3. Hence, the probability of winning is 0.

* Suppose the car was behind door 1 and the player chooses door 2. Then the host has no choice but to open door 3. Under the switching strategy, the player will definitely win. The car is behind door 1 with probability $\frac{1}{3}$ and the player chooses door 2 with probability $\frac{1}{3}$. Hence the overall probability of winning in this case is $\frac{1}{9}$. The same goes for when the car was behind door 1 and the player chooses door 3.

* So when the prize is in door 1, the probability of winning under the switching strategy is $\frac{1}{9} + \frac{1}{9} = \frac{2}{9}$. The prize could be behind any of the three doors; in each case it would be a repeat of the analysis done above. In conclusion, the total probability of winning under the switching strategy is $3 * \frac{2}{9} = \frac{2}{3}$. $\square$

Now consider the same problem with five doors instead of three. We carry out a similar analysis as the one above. We will assume throughout that the car is behind door 1, and multiply our answer by 5 to get the final probability.

* Again, if the player chooses door 1, there is no chance of winning by the switching strategy.

* If the player chooses door 2, then the host has a choice of opening doors 3, 4, or 5. Suppose he opens door 3. Then there is a $\frac{1}{3}$ chance of winning the car prize (choosing door 1 out of doors 1, 4, and 5).

  In summary, there is a $\frac{1}{5}$ chance that the player picks door 2, a $\frac{1}{3}$ chance that the host opens door 3, and a $\frac{1}{3}$ chance that the switching strategy picks the prize door 1. Thus, there is a total probability of $\frac{1}{5} * \frac{1}{3} * \frac{1}{3}$ of winning for this specific case.

* The same probability is derived for when the host opens door 4 or 5 instead. So we have a total winning probability of $3 * \frac{1}{5} * \frac{1}{3} * \frac{1}{3} = \frac{1}{5} * \frac{1}{3}$. Now this same probability is derived for when the player chooses to open door 3, 4, or 5 instead. So we have a total winning probability of $4 * \frac{1}{5} * \frac{1}{3}$.

* Finally, this same probability is derived for when the car prize is behind doors 2, 3, 4, or 5. The chance of each occurrence is $\frac{1}{5}$, which is not yet been factored into our analysis yet. So we have a total winning probability of $5 * 4 * \frac{1}{5} * \frac{1}{5} * \frac{1}{3} = 4 * \frac{1}{5} * \frac{1}{3} = \frac{4}{15}$. We have now covered all the cases so this is our final result. $\square$

Carrying out the same analysis, with n doors instead of five, will allow us to derive the following formula:

The chance of winning under the switching strategy is $\frac{n-1}{n(n-2)}$.

Now consider an extension of the problem where we still have n doors, but the host opens m goat doors, where $m < n - 1$.

For sake of ease, plug in some numerical values first, say $n = 6$ and $m = 2$. Suppose the prize is behind door 1.

* Note that the player can choose any of doors 1 to 6 with equal probability: $\frac{1}{6}$. Suppose the player picks door 2 (recall with door 1, there is no chance of winning under the switching strategy). Then the host can choose 2 out of the remaining 4 non-prize doors available with equal probability: $\frac{1}{\binom{4}{2}}$ Suppose the host chooses to open doors 3 and 4. Then under the switching strategy, the player can win with a probability of $\frac{1}{3}$. Hence, there is an overall probability of $\frac{1}{6} * \frac{1}{\binom{4}{2}} * \frac{1}{3}$ of winning with the switching strategy.

* There are $\binom{4}{2} = 6$ combinations of doors that the host can choose. And the above probability is the same for when the player chooses any of door 2, 3, 4, 5, 6. Furthermore, the prize door can also be any of the six doors, and it is with equal probability that each door can be the prize door. Hence, the total probability is $6 * 6 * 5\frac{1}{6} * \frac{1}{6} * \frac{1}{\binom{4}{2}} * \frac{1}{3} = \frac{5}{6*3} = \frac{5}{18}$. $\square$

Back to the generalized extension of n doors and having the host open m of the goat doors, a similar analysis to the above yields the following formula:

The chance of winning under the switching strategy is $\frac{n-1}{n(n-m-1)}$.

Now consider another extension of the problem where we have n doors and k of these doors have car prizes behind them, where $k < n - 1$.

For sake of ease, look at the case where $n = 10$ and $k = 3$ first. Suppose the prizes are behind doors 1, 2, and 3. This occurs with probability $\frac{1}{\binom{10}{3}}$ (it is only one of $\binom{10}{3}$ permutations of doors which the three prizes could be behind). But the following result will hold for all these door permutations. Hence, we also have a factor of $\binom{10}{3}$. So the two factors cancel each other out.

* Consider the case where the player chooses door 1. This occurs with probability $\frac{1}{10}$. There are then 7 remaining goat doors that the host can open. Suppose he opens door 4. This occurs with probability 1/7. Then the player has eight remaining doors to switch to but can only win by switching to either of the other two prize doors. Hence, the chance of winning

is $\frac{2}{8}$. This is the case for when the host opens doors 5 to 10 instead. So the probability of winning is $\frac{1}{10} * \frac{2}{8}$. This case repeats for if the player chooses door 2 or 3, since they are also prize doors. Hence, the total probability of winning is $3 * \frac{1}{10} * \frac{2}{8}$.

* Consider the case where the player chooses door 4. This occurs with probability $\frac{1}{10}$. There are then 6 remaining goat doors that the host can open. Suppose he opens door 5. This occurs with probability $1/6$. Then the player has eight remaining doors to switch to but can only win by switching to the three prize doors. Hence, the chance of winning is $\frac{3}{8}$. This is the case for when the host opens doors 6 to 10 instead. So the probability of winning is $\frac{1}{10} * \frac{3}{8}$. This case repeats for if the player chooses one of doors 5 to 10 instead, since they are also goat doors. Hence, the total probability of winning is $7 * \frac{1}{10} * \frac{3}{8}$.

* Thus, the collective probability of winning is $\frac{1}{10}\left[3 * \frac{2}{8} + 7 * \frac{3}{8}\right]$. $\square$

Back to the generalized extension of n doors and having k of these doors hold prizes, a similar analysis to the above yields the following formula:

The chance of winning under the switching strategy is $\frac{1}{n}\left[k * \frac{k-1}{n-2} + (n-k) * \frac{k}{n-2}\right]$.

Finally, consider the extension where everything is combined together: we have n doors, k of which have prizes behind them, and the host opens m doors, where $m < k+1$ and $k < n-1$. What is the chance of winning a prize under the switching strategy?

For sake of ease, look at the case where $n = 10$, $m = 2$, and $k = 3$ first. Suppose the prizes are behind doors 1, 2, and 3. This occurs with probability $\frac{1}{\binom{10}{3}}$ (it is only one of $\binom{10}{3}$ permutations of doors which the three prizes could be behind). But the following result will hold for all these door permutations. Hence, we also have a factor of $\binom{10}{3}$. So the two factors cancel each other out.

* Consider the case where the player chooses door 1. This occurs with probability $\frac{1}{10}$. There are then 7 remaining goat doors that the host can open. Suppose he opens door 4 and door 5. This occurs with probability $\frac{1}{\binom{7}{2}}$. Then the player has seven remaining doors to switch to but can only win by switching to either of the other two prize doors. Hence, the chance of winning is $\frac{2}{7}$. This is the case for when the host opens any other pair of goat doors instead. So the probability of winning is $\frac{1}{10} * \frac{2}{7}$. This case repeats for if the player chooses door 2 or 3, since they are also prize doors. Hence, the total probability of winning is $3 * \frac{1}{10} * \frac{2}{7}$.

* Consider the case where the player chooses door 4. This occurs with probability $\frac{1}{10}$. There are then 6 remaining goat doors that the host can open. Suppose he opens door 5 and door 6. This occurs with probability $\frac{1}{\binom{6}{2}}$. Then the player has seven remaining doors to switch to but can only win by switching to the three prize doors. Hence, the chance of winning is $\frac{3}{7}$. This is the case for when the host opens any other pair of goat doors (excluding the door that the player selected) instead. So the probability of winning is $\frac{1}{10} * \frac{3}{7}$. This case repeats for if the player chooses one of doors 5 to 10 instead, since they are also goat doors. Hence, the total probability of winning is $7 * \frac{1}{10} * \frac{3}{7}$.

* Thus, the collective probability of winning is $\frac{1}{10}\left[3 * \frac{2}{7} + 7 * \frac{3}{7}\right]$. $\square$

chekc calculations

Back to the generalized combined extension of n doors, having k of these doors hold prizes and having the host open m goat doors, a similar analysis to the above yields the following formula:

The chance of winning under the switching strategy is $\frac{1}{n}\left[k * \frac{k-1}{n-m-1} + (n-k) * \frac{k}{n-m-1}\right]$.

different value of prizes behind each door extension to problem

Gambler's Ruin Problem

Suppose we play a game where the gambler starts out with 90 dollars and bets one dollar per turn until he gains an additional 10 dollars (i.e, 'wins the game') or he loses all of his money (i.e, 'loses the game'). At each turn he chooses a color on the roulette wheel and spins it. The probability of winning a turn is $\frac{9}{19}$. We want to determine the probability of the gambler winning the overall game before he loses all his money (goes to ruin).

Let W denote the event that you win the game (get 10 more dollars before lose all 90 dollars) and let D denote the amount of money you have at your current time. Further denote

$$p_i := P(W|D = i)$$

The quantity we desire is then p_{90}.

Consider the various values that p_i can be.

1. $i = 0$ means that you have no money. This means that you cannot even play the game, so your chance of winning is 0.

2. $i = 90 + 10 = 100$ means that you've already reached your goal amount of money. Your chance of winning is then 1.

3. For any integer value of i between 0 and 100, there are two scenarios. Losing your current turn occurs with probability $\frac{10}{19}$. Once you lose your current turn, you end up with one less dollar than you had before: $i - 1$. From there, the probability of winning the game is p_{i-1}. On the other hand, if you win your current turn, you gain a dollar and your probability of winning becomes p_{i+1}. Hence, the probability p_i is:

$$p_i = \frac{10}{19} * p_{i-1} + \frac{9}{19} * p_{i+1}$$

Overall,

$$p_i = \begin{cases} 0 \text{ if } i = 0 \\ 1 \text{ if } i = 100 \\ \frac{10}{19} * p_{i-1} + \frac{9}{19} * p_{i+1} \text{ else} \end{cases}$$

Rewrite the last case as

$$\begin{aligned} p_i &= \frac{10}{19} * p_{i-1} + \frac{9}{19} * p_{i+1} \\ \frac{10}{19} * p_i + \frac{9}{19} * p_i &= \frac{10}{19} * p_{i-1} + \frac{9}{19} * p_{i+1} \\ \frac{10}{19}(p_i - p_{i-1}) &= \frac{9}{19}(p_{i+1} - p_i) \\ p_{i+1} - p_i &= \frac{10}{9}(p_i - p_{i-1}) \end{aligned}$$

Plug in $i = 1, 2, 3, \cdots$ into the recursive relationship above to obtain the series of equations:

$$
\begin{aligned}
p_2 - p_1 &= \frac{10}{9}(p_1 - p_0) = \frac{10}{9}p_1 \\
p_3 - p_2 &= \frac{10}{9}(p_2 - p_1) = \left(\frac{10}{9}\right)^2 p_1 \\
p_4 - p_3 &= \frac{10}{9}(p_3 - p_2) = \left(\frac{10}{9}\right)^3 p_1 \\
&1\vdots \\
p_{i+1} - p_i &= \frac{10}{9}(p_i - p_{i-1}) = \left(\frac{10}{9}\right)^i p_1
\end{aligned}
$$

Summing up the first $i-1$ of these equations yields:

$$
\begin{aligned}
p_i - p_1 &= p_1 \left[\frac{10}{9} + \left(\frac{10}{9}\right)^2 \left(\frac{10}{9}\right)^3 + \cdots \left(\frac{10}{9}\right)^{i-1}\right] \\
p_i &= p_1 \left[1 + \frac{10}{9} + \left(\frac{10}{9}\right)^2 \left(\frac{10}{9}\right)^3 + \cdots \left(\frac{10}{9}\right)^{i-1}\right] \\
&= p_1 \left[\frac{1 - \left(\frac{10}{9}\right)^i}{1 - \frac{10}{9}}\right]
\end{aligned}
$$

We can solve for p_1 by substituting $i = 100$ because we already know the value of p_{100}:

$$
\begin{aligned}
p_{100} &= p_1 \left[\frac{1 - \left(\frac{10}{9}\right)^{100}}{1 - \frac{10}{9}}\right] \\
1 &= p_1 \left[\frac{1 - \left(\frac{10}{9}\right)^{100}}{1 - \frac{10}{9}}\right] \\
p_1 &= \frac{1 - \frac{10}{9}}{1 - \left(\frac{10}{9}\right)^{100}}
\end{aligned}
$$

The value we desire is p_{90}. Plugging $i = 90$ into the equation:

$$
\begin{aligned}
p_{90} &= p_1 \left[\frac{1 - \left(\frac{10}{9}\right)^{90}}{1 - \frac{10}{9}}\right] \\
&= \frac{1 - \left(\frac{10}{9}\right)^{90}}{1 - \left(\frac{10}{9}\right)^{100}}
\end{aligned}
$$

Now we can generalize our above calculations. Suppose we play a game where the gambler starts out with n and bets a portion of this money per turn until he gains an additional $m < n$ dollars or he loses all his money. Further suppose the probability of winning a bet per turn is $p < \frac{1}{2}$. We again want to determine the probability of winning the overall game before the gambler goes to ruin.

For this problem, we will consider several different betting strategies. We will show that the strategy of betting one dollar each turn will almost always be likely to result in losing all n dollars before gaining m dollars, no matter how much money you start off with.

Let W be the event that we reach a total sum of $T = n + m$ dollars before we lose all n dollars, and let D be the amount of money you start with. Then we are interested in computing $P(W|D = n)$.

Let $p_i = P(W|\{D = i\})$. Then we claim that

$$p_i = \begin{cases} 0 \text{ if } i = 0 \\ 1 \text{ if } i = T \\ p * p_{i+1} + (1-p) * p_{i-1} \text{ if } 0 < i < T \end{cases}$$

To prove this, we will first see that the first two cases are obvious: if $n = 0$, there is no chance of reaching T total dollars and if $i = T$, then we've reached our goal already.

To show the last case, we will first define E to be the event we win the first bet and $\overline{E}$ to be otherwise. Then

$$\begin{aligned} p_i &= P(W, E|\{D = i\}) + P(W, \overline{E}|\{D = i\}) \\ &= P(W|E, \{D = i\})P(E|\{D = i\}) + P(W|\overline{E}, \{D = i\})P(\overline{E}|\{D = i\}) \\ &= p * p_{i+1} + (1-p) * p_{i-1} \end{aligned}$$

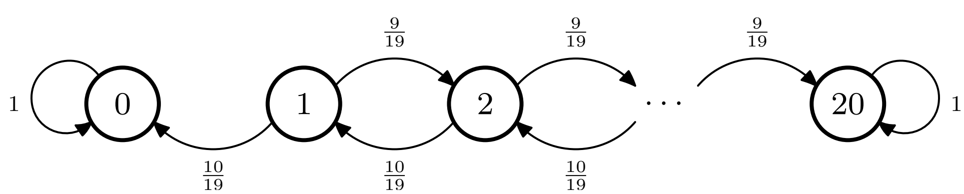
Solving this linear, homogeneous difference equation with the boundary conditions specified by the other two cases as before, we get that

$$p_n = \frac{\left(\frac{1-p}{p}\right)^n - 1}{\left(\frac{1-p}{p}\right)^T - 1}$$

Going back to the numerical setup of the problem, consider the case where we instead bet 2 dollars instead of 1 at each turn. It would take half as many steps to reach our goal of 100 dollars because we are betting twice as much. So we can scale the whole picture down by a factor of 2, that is, it is the same as using the one dollar per bet strategy when $m = 5, n = 45, p = \frac{9}{19}$.

As an example, suppose we bet 5 dollars at a time. Then we would have to scale everything down by a factor of 5 and perform computations for the case where we bet one dollar per turn when $m = 2, n = 18$ and $p = \frac{9}{19}$.

We then have the following Markov Chain:



Once again, we are interested in the probability of reaching our goal of 20 dollars starting from state 18, i.e, p_{18}. We set up the following probability equations:

$$\begin{aligned}
p_0 &= 0 \\
p_1 &= \frac{9}{19}p_2 \\
p_2 &= \frac{10}{19}p_1 + \frac{9}{19}p_3 \\
p_3 &= \frac{10}{19}p_2 + \frac{9}{19}p_4 \\
p_4 &= \frac{10}{19}p_3 + \frac{9}{19}p_5 \\
&\vdots \\
p_{18} &= \frac{10}{19}p_{17} + \frac{9}{19}p_{19} \\
p_{19} &= \frac{9}{19} + \frac{10}{19}p_{18} \\
p_{20} &= 1
\end{aligned}$$

Solving this system of equations, we get that p_{18} is 0.7837. There is a 78.37 percent chance of reaching our desired goal of 20 dollars using the one dollar per bet strategy if we start from 18 dollars. (Appendix Figure 1)

Now suppose we are interested in the mean time it takes to end the game, whether you get your prize or you lose all your money, i.e, the mean time it takes to get absorbed by either state 0 or state 20.

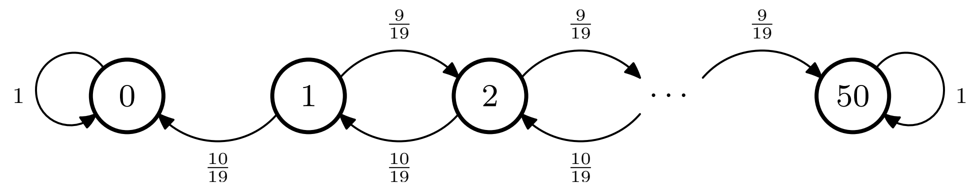
We set up the following probability equations:

$$\begin{aligned}
m_0 &= 0 \\
m_1 &= \frac{10}{19}(1) + \frac{9}{19}(1 + m_2) \\
m_2 &= \frac{10}{19}(1 + m_1) + \frac{9}{19}(1 + m_3) \\
m_3 &= \frac{10}{19}(1 + m_2) + \frac{9}{19}(1 + m_4) \\
m_4 &= \frac{10}{19}(1 + m_3) + \frac{9}{19}(1 + m_5) \\
&\vdots \\
m_{18} &= \frac{10}{19}(1 + m_{17}) + \frac{9}{19}(1 + m_{19}) \\
m_{19} &= \frac{9}{19}(1) + \frac{10}{19}(1 + m_{18}) \\
m_{20} &= 0
\end{aligned}$$

Solving this system of equations, we get that m_{18} is 44.1927. It takes a mean time of 44 time steps to get finish the game if we use the one dollar per bet strategy starting from 18 dollars. (Appendix Figure 2)

Suppose we instead bet 20 dollars at a time, so that five good turns from a start of 900 dollars will enable us to reach our goal of 1000 dollars. Then we would have to scale everything down by a factor of 20 and perform computations for the case where we bet one dollar per turn when $m = 5, n = 45$ and $p = \frac{9}{19}$.

We then have the following Markov Chain:



Once again, we are interested in the probability of reaching our goal of 20 dollars, starting from every state. We set up the following probability equations:

$$
\begin{aligned}
p_0 &= 0 \\
p_1 &= \frac{9}{19}p_2 \\
p_2 &= \frac{10}{19}p_1 + \frac{9}{19}p_3 \\
p_3 &= \frac{10}{19}p_2 + \frac{9}{19}p_4 \\
p_4 &= \frac{10}{19}p_3 + \frac{9}{19}p_5 \\
&\vdots \\
p_{48} &= \frac{10}{19}p_{47} + \frac{9}{19}p_{49} \\
p_{49} &= \frac{9}{19} + \frac{10}{19}p_{48} \\
p_{50} &= 1
\end{aligned}
$$

Solving this system of equations, we get that p_{18} is 0.5884. (Appendix Figure 3) There is a 58.84 percent chance of reaching our desired goal of 20 dollars using the one dollar per bet strategy if we start from 18 dollars. The chance of winning is noticeably slimmer than when we bet 50 dollars per turn.

Now again suppose we are interested in the mean time it takes to end the game, whether you get your prize or you lose all your money, i.e, the mean time it takes to get absorbed by either state 0 or state 50.

We set up the following probability equations:

$$
\begin{aligned}
m_0 &= 0 \\
m_1 &= \frac{10}{19}(1) + \frac{9}{19}(1 + m_2) \\
m_2 &= \frac{10}{19}(1 + m_1) + \frac{9}{19}(1 + m_3) \\
m_3 &= \frac{10}{19}(1 + m_2) + \frac{9}{19}(1 + m_4) \\
m_4 &= \frac{10}{19}(1 + m_3) + \frac{9}{19}(1 + m_5) \\
&\vdots \\
m_{48} &= \frac{10}{19}(1 + m_{47}) + \frac{9}{19}(1 + m_{49}) \\
m_{49} &= \frac{9}{19}(1) + \frac{10}{19}(1 + m_{48}) \\
m_{50} &= 0
\end{aligned}
$$

We get that m_{45} is 296.0499, so it takes a mean time of 296 time steps to get finish the game if we use the one dollar per bet strategy starting from 45 dollars (Appendix Figure 4). The duration of the game is noticeably longer than when we bet 50 dollars per turn.

Best Strategy:

Intuitively, the best strategy would be to bet m dollars at once, as the probability of reaching the goal would be significantly larger (probability p of winning as opposed to p^m, in the case of the betting one dollar per game strategy). We will prove this claim more rigorously.

Suppose again that $m = 1, n = 9$, and $p = \frac{9}{19}$ (so that $q = \frac{10}{19}$). In the first game, we will bet 1 dollar so that the chance of reaching our goal of 10 dollars is $\frac{9}{19}$. And if we lose, we are left with 8 dollars. So

$$
\begin{aligned}
p_9 &= \frac{9}{19} + \frac{10}{19}p_8 \\
p_9 - \frac{10}{19}p_8 &= \frac{9}{19}
\end{aligned}
$$

Now from 8 dollars, we will bet 2 dollars so that again, the probability of winning is $\frac{9}{19}$ and we end up with 6 dollars left with probability $\frac{10}{19}$. Hence

$$p_8 - \frac{10}{19}p_6 = \frac{9}{19}$$

Repeating again for while we start at 6 dollars, we get:

$$p_6 - \frac{10}{19}p_2 = \frac{9}{19}$$

Now we bet all of our money, since no proper subset amount of it will enable us to reach our goal.

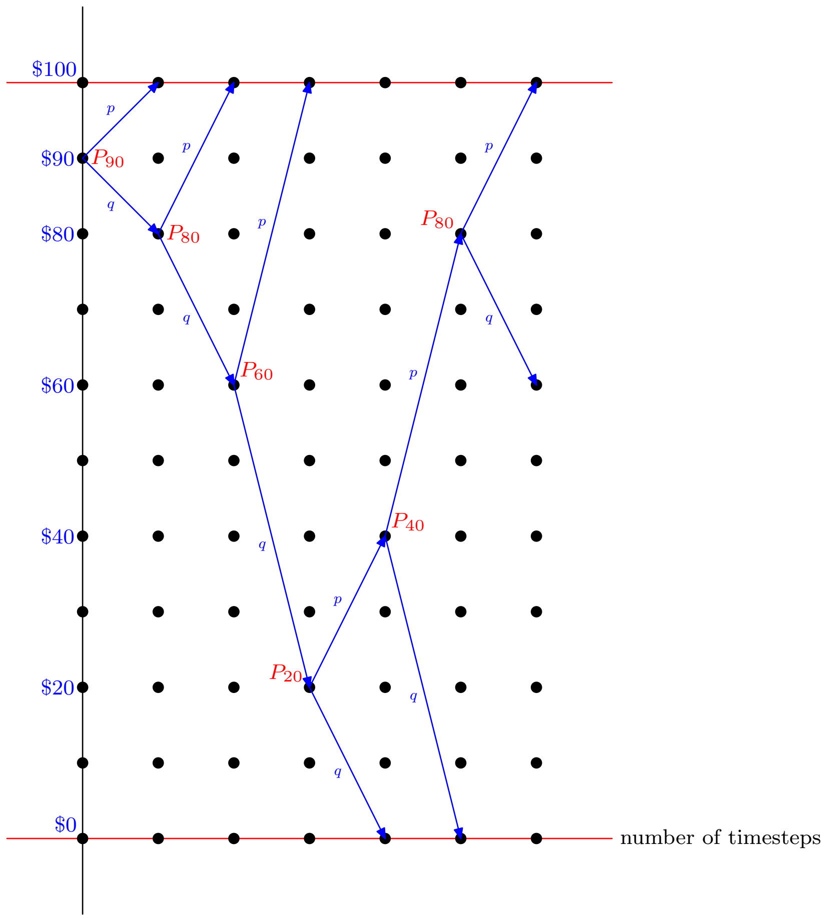
So we get:

$$p_2 = \frac{9}{19}p_4 + \frac{10}{19}p_0$$
$$= \frac{9}{19}p_4$$
$$p_2 - \frac{9}{19}p_4 = 0$$

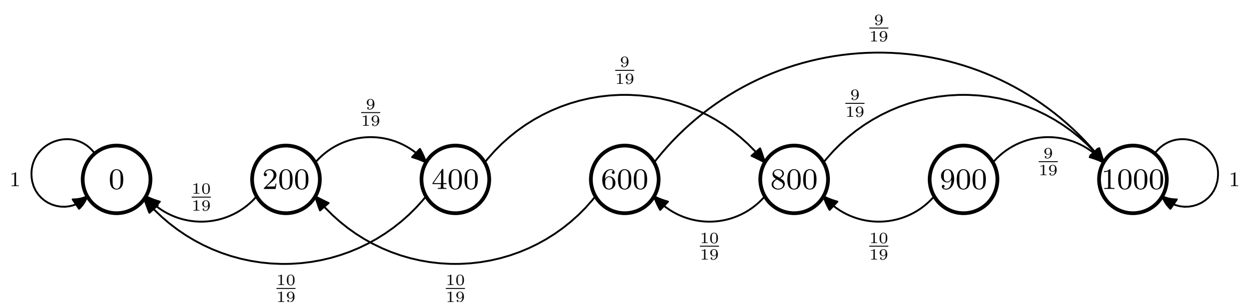
Similarly for while we have 4 dollars in our pocket:

$$p_4 - \frac{9}{19}p_8 = 0$$

Overall, we have the following diagram:



with the corresponding Markov Chain:



In matrix representation, we have

$$\begin{bmatrix} 1 & -\frac{10}{19} & 0 & 0 & 0 \\ 0 & 1 & -\frac{10}{19} & 0 & 0 \\ 0 & 0 & 1 & 0 & -\frac{10}{19} \\ 0 & 0 & 0 & -\frac{9}{19} & 1 \\ 0 & -\frac{9}{19} & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} p_9 \\ p_8 \\ p_6 \\ p_4 \\ p_2 \end{bmatrix} = \begin{bmatrix} \frac{9}{19} \\ \frac{9}{19} \\ \frac{9}{19} \\ 0 \\ 0 \end{bmatrix}$$

Solving this matrix equation, we get:

$$\begin{bmatrix} p_9 \\ p_8 \\ p_6 \\ p_4 \\ p_2 \end{bmatrix} = \begin{bmatrix} 0.8794 \\ 0.7709 \\ 0.5647 \\ 0.3652 \\ 0.1730 \end{bmatrix}$$

And so, there is a 0.8794 chance of going home with 10 dollars before losing all your money. Note that in the symmetrical case ($p = \frac{1}{2}$), the chance of winning 10 dollars, starting with 9 dollars, is $\frac{9}{10}$. Here, because the probability of winning each round is slightly less than a half, we expect that the probability of winning the whole game is also slightly less than $\frac{9}{10}$. So our result is reasonable.

Aside:

A few computational problems arise in the study of Markov Chains. We will consider the following two:

1. How many times will a transient state be occupied before absorption occurs?

   Let e_{ij} denote the mean number of times that transient state j is occupied given that the initial state (also transient) is i.

   We get that if $e_{ij} = \sum\limits_{\substack{k \\ \text{transient}}} p_{ik} e_{kj}$ if $i \neq j$ and that $e_{ii} = 1 + \sum\limits_{\substack{k \\ \text{transient}}} p_{ik} e_{kj}$ if $i = j$.

   To obtain a more generalized form, we will use matrices. The above equations can be written as:

   $$E = I + QE$$
   $$\implies E = (I - Q)^{-1}$$

   where Q is the matrix of transition probabilities of transient states to another transient state and E is the matrix of all e_{ij}'s.

2. Suppose the absorbing state is specified. How many transitions will occur before the specified state is occupied?

   Intuitively, is there is more than one absorbing state, we don't know where the final transition will occur. This implies that the mean first passage time may be infinite.

Hence, we define m_{ij} to be the conditional mean first passage time from i to j. To further account for this uncertainty, we multiply each m_{ij} with their associated absorbing probabilities a_{ij}.

We obtain equations of the following form, for each i, j:

$$a_{ij} = 1 * a_{ij} + \sum_{\substack{k \\ \text{transient}}} p_{ik} a_{kj} m_{kj}$$

where the first term comes from the possibility that in one step, we get absorbed into our desired absorbing state, and the second term comes from the possibility that we don't transition to j right away and instead traverse to some middle state k.

We will now proceed to answer question two as above for our best strategy example: suppose we wanted to calculate the probability of getting absorbed into state j, where $j \in \{0, 10\}$, starting from state $i \in \{2, 4, 6, 8, 9\}$. Let $a_{i,j}$ denote the probability of reaching state j from state i. Then we have the general expression:

$$a_{i,j} = a_{i,j} + \sum_{\substack{k \\ \text{transient}}} p_{i,k} a_{k,j}$$

where the first term is the probability of reaching j from i in one step, and the second term sums the probabilities of reaching j from i in more than two steps.

Furthermore, we can see that $a_{0,10} = a_{10,0} = 0$ since both are absorbing states, so there is no way to reach one from the other.

Plugging in the correct values into the expression above for all i, j, we construct the following equations:

$$a_{2,0} = \frac{10}{19} + \frac{9}{19} a_{4,0}$$
$$a_{4,0} = \frac{10}{19} + \frac{9}{19} a_{8,0}$$
$$a_{6,0} = 0 + \frac{10}{19} a_{2,0}$$
$$a_{8,0} = 0 + \frac{10}{19} a_{6,0}$$
$$a_{9,0} = 0 + \frac{10}{19} a_{8,0}$$
$$a_{2,10} = 0 + \frac{9}{19} a_{4,10}$$
$$a_{4,10} = 0 + \frac{9}{19} a_{8,10}$$
$$a_{6,10} = \frac{9}{19} + \frac{10}{19} a_{2,10}$$
$$a_{8,10} = \frac{9}{19} + \frac{10}{19} a_{6,10}$$
$$a_{6,10} = \frac{9}{19} + \frac{10}{19} a_{8,10}$$

All the calculated numerical values are summarized in Appendix Figure 5. For example, the probability of reaching state 0 from state 2 ($a_{2,0}$) is 82.7%.

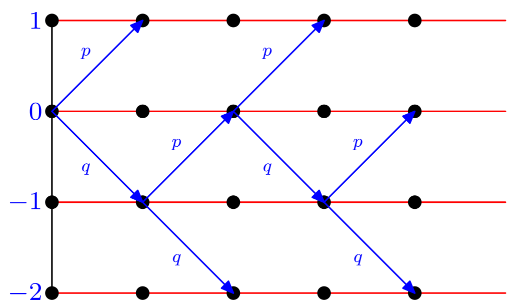
Now, we calculate the expected number of games that must be played in order to win, using this scenario. Let S be the number of games played until we either win our desired amount or we lose everything we have.

Then, similar to the probabilities, we have the following:

$$m_{i/100} = E\left[S|D=i\right] = \begin{cases} 0 \text{ if } i=0 \\ 0 \text{ if } i=T \\ 1+pE\left[S|D=i+1\right]+(1-p)E\left[S|D=i-1\right] \text{ if } 0<i<T \end{cases}$$

The above claim can be proven in much the same way as previously done.

To get some intuition, let us consider the following toy example:



We can compute this in two ways.

First, let m denote the mean time it takes to get absorbed into either state -2 or 1. Then we set up the usual mean first passage time equations and solve for m.

$$\begin{aligned} m &= \frac{1}{2}(1) + \frac{1}{4}(2+m) + \frac{1}{4}(2) \\ &= \frac{1}{2} + \frac{1}{4}m + \frac{1}{2} + \frac{1}{2} \\ &= \frac{3}{2} + \frac{1}{4}m \\ \implies \frac{3}{4}m &= \frac{3}{2} \end{aligned}$$

Hence we get that $m = 2$.

Or we can brute force the entire calculation by using the definition of expected value and multiply the probability and the length of the sample path for all sample paths that lead to absorption:

$$m = \frac{1}{2}(1) + \frac{1}{2}^{2}(2) + \frac{1}{2}^{3}(3) + \cdots$$

Hence we also end up with $m = 2$.

Going back to our original problem, we set up equations similar to the ones constructed for probability above. Starting from 9 dollars, it takes us one time step to reach our goal of 10 dollars, but this occurs with probability $\frac{9}{19}$. Otherwise, we lose a dollar and have $1 + m_8$ more time steps to go. This occurs with probability $\frac{10}{19}$.

$$\begin{aligned} m_9 &= \frac{9}{19} * 1 + \frac{10}{19}(1 + m_8)) \\ &= 1 + \frac{10}{19} m_8 \\ m_9 - \frac{10}{19} m_8 &= 1 \end{aligned}$$

Similarly, starting from 8 dollars, we have:

$$m_8 - \frac{10}{19} m_6 = 1$$

From 6 dollars:

$$m_6 - \frac{10}{19} m_2 = 1$$

From 2 dollars, we bet all our money in one time step and if we lose (with probability $\frac{10}{19}$) we have no choice but to stop. Otherwise, we have 4 total dollars with probability $\frac{9}{19}$ and take $1 + m_4$ time steps. This occurs with probability $\frac{9}{19}$

$$m_2 - \frac{9}{19} m_4 = 1$$

Similarly, starting from 4 dollars we have:

$$m_4 - \frac{9}{19} m_8 = 1$$

Overall, in matrix representation, we get

$$\begin{bmatrix} 1 & -\frac{10}{19} & 0 & 0 & 0 \\ 0 & 1 & -\frac{10}{19} & 0 & 0 \\ 0 & 0 & 1 & 0 & -\frac{10}{19} \\ 0 & 0 & 0 & -\frac{9}{19} & 1 \\ 0 & -\frac{9}{19} & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} m_9 \\ m_8 \\ m_6 \\ m_4 \\ m_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

We get $m_9 = 2.0857, m_8 = 2.0627, m_6 = 2.0192, m_4 = 1.9771$, and $m_2 = 1.9365$.

Now consider the case where instead of an absorbing boundary at $D = 0$, we have a "reflecting boundary" there (i.e, you are immediately given 200 extra dollars if you run out of money). Suppose again we have $n = 900, m = 100$ but we play a fair game so that $p = q = \frac{1}{2}$. Because there is no way of losing the game, the probability of winning is 1 no matter how much you start out with.

We set up the following expected value equations. Note that the equations for m_9, m_8, and m_6 are the same because they never reach 0 dollars.

$$\begin{aligned} m_9 - \frac{1}{2} m_8 &= 1 \\ m_8 - \frac{1}{2} m_6 &= 1 \\ m_6 - \frac{1}{2} m_2 &= 1 \end{aligned}$$

When we are at 2 dollars, with probability $\frac{1}{2}$ we end up with 4 dollars, from which it takes mean time m_4 to get to 10 dollars. With probability $\frac{1}{2}$, we lose all our money, and we end up borrowing 2 dollars again from someone else, from which it takes mean time m_2 to get to 10 dollars. Let us count that as two timesteps.

$$\begin{aligned} m_2 &= \frac{1}{2}(1+m_4) + \frac{1}{2}(2+m_2) \\ \frac{1}{2}m_2 - \frac{1}{2}m_4 &= \frac{3}{2} \\ m_2 - m_4 &= 3 \end{aligned}$$

Similarly, when we are at 4 dollars and we bet all of it in one turn:

$$\begin{aligned} m_4 &= \frac{1}{2}(1+m_8) + \frac{1}{2}(2+m_2) \\ m_4 - \frac{1}{2}m_8 - \frac{1}{2}m_2 &= \frac{3}{2} \end{aligned}$$

Altogether, we have the following matrix equation:

$$\begin{bmatrix} 1 & -\frac{1}{2} & 0 & 0 & 0 \\ 0 & 1 & -\frac{1}{2} & 0 & 0 \\ 0 & 0 & 1 & 0 & -\frac{1}{2} \\ 0 & 0 & 0 & -1 & 1 \\ 0 & -\frac{1}{2} & 0 & 1 & -\frac{1}{2} \end{bmatrix} \begin{bmatrix} m_9 \\ m_8 \\ m_6 \\ m_4 \\ m_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 3 \\ \frac{3}{2} \end{bmatrix}$$

We obtain $m_9 = 3.5, m_8 = 5, m_6 = 8, m_4 = 11, m_2 = 14$.

Now suppose we have a specified absorbing state. We wish to determine how many transitions it will take before the specified state is reached.

Let m_{ij} represent the conditional mean first passage time from i to j. Then we get:

$$a_{ij}m_{ij} = a_{ij} + \sum_{k \neq i} p_{ik}a_{kj}m_{kj}$$

where the first term denotes the transition in one step and the second term is the sum of all transitions that take more than one step.

Plugging in the correct values into the expression above for all i, j, we construct the following equations. Each value of $a_{i,j}$ has been computed above.

$$\begin{aligned}
a_{9,10}m_{9,10} &= a_{9,10} + p_{9,8}a_{8,10}m_{8,10}\\
a_{8,10}m_{8,10} &= a_{8,10} + p_{8,6}a_{6,10}m_{6,10}\\
a_{6,10}m_{6,10} &= a_{6,10} + p_{6,2}a_{2,10}m_{2,10}\\
a_{4,10}m_{4,10} &= p_{4,8}a_{8,10}m_{8,10}\\
a_{2,10}m_{2,10} &= p_{2,4}a_{4,10}m_{4,10}\\
a_{9,0}m_{9,0} &= p_{9,8}a_{8,0}m_{8,0}\\
a_{8,0}m_{8,0} &= p_{8,6}a_{6,0}m_{6,0}\\
a_{6,0}m_{6,0} &= p_{6,2}a_{2,0}m_{2,0}\\
a_{4,0}m_{4,0} &= a_{4,0} + p_{4,8}a_{8,0}m_{8,0}\\
a_{2,0}m_{2,0} &= a_{2,0} + p_{2,4}a_{4,0}m_{4,0}
\end{aligned}$$

All the calculated numerical values are summarized in Appendix Figure 6. For example, the expected number of timesteps taken to reach state 0, starting from state 2 is 1.7111.

Additional Extensions to Gambler's Ruin

Extension 1:

Suppose we gamble using the following strategy: we start out with 8 dollars in our pocket and bet one dollar per turn. If we reach our desired goal of 10 dollars then we play the game again starting with 9 dollars out of the 10 dollars we now have. If we win again, we play the game again starting with 8 dollars out of the 11 dollars we now have. This continues until the first game where we lose all of the 8 dollars we start out with. Winning a turn has probability $\frac{9}{19}$.

We wish to determine the expected amount of money you end up with after the first time you lose. We also want to determine the expected number of turns taken until you lose.

We've seen the following formula in previous renditions of the problem:

$$p_i = \begin{cases} 0 \text{ if } i = 0 \\ 1 \text{ if } i = 10 \\ \frac{9}{19} * p_{i+1} + \frac{10}{19} * p_{i-1} \text{ if } 0 < i < 10 \end{cases}$$

where p_i is the probability of winning a game, starting with i dollars.

Now we generalize the above calculations. We start out with n dollars in our pocket and bet one dollar per turn. If we reach our desired goal of T dollars then we play the game again starting with n dollars out of the T dollars we now have. If we win again, we play the game again starting with n dollars out of whatever total amount of money we now have. This continues until the first game where we lose all of the n dollars we start out with.

We ask the same question as above: determine the expected amount of money you end up with after you lose.

Winning a turn (gaining a dollar) has probability p and losing has probability $1 - p$. We've seen in the very first rendition of the Gambler's Ruin Problem that:

$$p_i = \begin{cases} 0 \text{ if } i = 0 \\ 1 \text{ if } i = T \\ p * p_{i+1} + (1-p) * p_{i-1} \text{ if } 0 < i < T \end{cases}$$

where p_i is the probability of winning a game, starting with i dollars.

We can draw out the following tree. The number at each node denotes the amount of money you have in your pocket at that time and the number at each edge is the probability of reaching that sum of money.

Define W_n to be the random variable denoting the total amount of money you end up with, starting this strategy from initial amount of money n. We can compute the expected value of W (which is the quantity we desire) by writing out the sum:

$$\begin{aligned} E[W] &= p_n E[W_T] + (1-p_n)(0) \\ &= {p_n}^2 E[W_{T+m}] + p_n(1-p_n)(m) + (1-p_n)(0) \\ &= {p_n}^3 E[W_{T+2m}] + {p_n}^2(1-p_n)(2m) + p_n(1-p_n)(m) + (1-p_n)(0) \\ &= \cdots \end{aligned}$$

Now we can solve for this expectation by using conditional expectations. Let k be the number of games played until we lose. Note that $E[k] = \frac{1}{1-p_n}$ because we have a geometric series where the probability of obtaining the losing outcome is $1-p_n$.

So:

$$\begin{aligned} E[W] &= E[E[W|k]] \\ &= {p_n}^k E\left[W_{T+(k+1)m}\right] + \cdots + {p_n}^2(1-p_n)(2m) + p_n(1-p_n)(m) + (1-p_n)(0) \end{aligned}$$

Extension 2:

Consider a *similar* setup as Extension 1. We gamble by starting out with 9 dollars in our pocket and bet using the *best strategy detailed previously.* If we reach our desired goal of 10 dollars then we play the game again starting with 9 out of the 10 dollars we now have, keeping one dollar to ourselves. If we win again, we play the game again starting with 9 dollars out of the 11 dollars we now have, keeping 2 dollars to ourselves. This continues until the first game we lose all of the 9 dollars we start out with.

We want to determine the probability that we've amassed 20 dollars before we lose the game. Furthermore, we want to determine the expected number of turns taken until we've amassed 20 dollars.

Extension 3:

Suppose now we play the above game but instead of starting with n after the first game, we start with $n + \sum_{i=1}^{k} \lceil \frac{m}{2^i} \rceil$ where $m = T - n$ and k is the number of games played after the first one.

We wish to compute the probability of reaching the point where we start betting T dollars per game before ruin.

We will first consider a numerical example where $T = 100$ and $n = 60$. Then the winning amount of money you start each game with until you win is $60 \to 60 + \frac{40}{2} = 80 \to 80 + \frac{20}{2} = 90 \to 95 \to 98 \to 99 \to 100$. A diagram of this possible play is shown in the figure below:

The only way to win the entire run is to win at each game. The probability of winning each game can be computed from the above expression of winning a game starting from n. Hence, we obtain the following expression:

$$\left[\frac{\left(\frac{1-p}{p}\right)^{60} - 1}{\left(\frac{1-p}{p}\right)^{100} - 1}\right] * \left[\frac{\left(\frac{1-p}{p}\right)^{80} - 1}{\left(\frac{1-p}{p}\right)^{100} - 1}\right] * \left[\frac{\left(\frac{1-p}{p}\right)^{90} - 1}{\left(\frac{1-p}{p}\right)^{100} - 1}\right] * \left[\frac{\left(\frac{1-p}{p}\right)^{95} - 1}{\left(\frac{1-p}{p}\right)^{100} - 1}\right] * \left[\frac{\left(\frac{1-p}{p}\right)^{98} - 1}{\left(\frac{1-p}{p}\right)^{100} - 1}\right] * p$$

With $p = \frac{1}{2}$, this expression evaluates to...

Now we will try this calculation for $n = 40$. The path to ultimate win is $40 \to 70 \to 85 \to 96 \to 98 \to 99 \to 100$. The probability expression is below:

$$\left[\frac{\left(\frac{1-p}{p}\right)^{40} - 1}{\left(\frac{1-p}{p}\right)^{100} - 1}\right] * \left[\frac{\left(\frac{1-p}{p}\right)^{70} - 1}{\left(\frac{1-p}{p}\right)^{100} - 1}\right] * \left[\frac{\left(\frac{1-p}{p}\right)^{85} - 1}{\left(\frac{1-p}{p}\right)^{100} - 1}\right] * \left[\frac{\left(\frac{1-p}{p}\right)^{96} - 1}{\left(\frac{1-p}{p}\right)^{100} - 1}\right] * \left[\frac{\left(\frac{1-p}{p}\right)^{98} - 1}{\left(\frac{1-p}{p}\right)^{100} - 1}\right] * p$$

regardless of where you start, it takes the same number of steps?? Check...

Now we will generalize this problem to a formula. We start at n dollars and climb our way up to the top.

$$\left[\frac{\left(\frac{1-p}{p}\right)^{n} - 1}{\left(\frac{1-p}{p}\right)^{100} - 1}\right] * \left[\frac{\left(\frac{1-p}{p}\right)^{n+\frac{m}{2}} - 1}{\left(\frac{1-p}{p}\right)^{100} - 1}\right] * \cdots * p$$

Now consider this game but instead of betting one dollar at a time starting from n, we bet $m := T - n$ dollars per turn, or if $m > n$, we bet all n dollars that we have.

Extension 4:

With reflecting boundary....

Sample Space Approach:

We start with the usual setup case: start with 9 dollars and bet one dollar per turn until we either reach 10 dollars or we lose all our money. You win a turn if the coin you flip comes out H and lose if T.

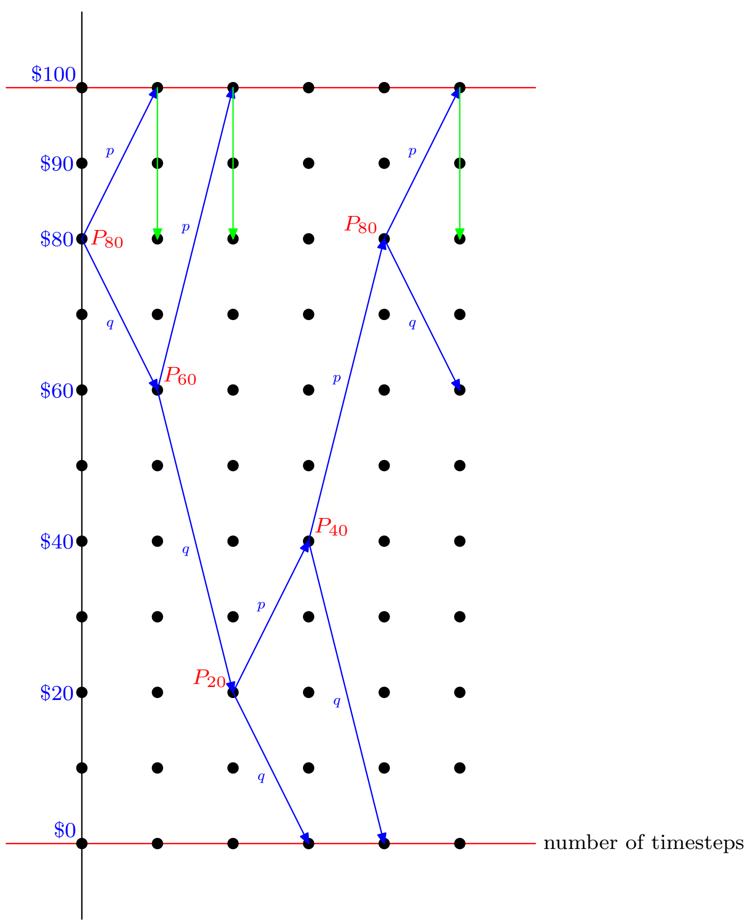
Let Ω_i denote the sample space at time i. Then at time $i = 1$, we either gain one dollar (and win the entire game) if we get H or we lose a dollar if we get T:

$$\Omega_1 = \{H, T\} = \{1, -1\}$$

Similarly for every time forward:

$$\begin{aligned}
\Omega_2 &= \{TH, TT\} = \{0, -2\} \\
\Omega_3 &= \{THH, THT, TTH, TTT\} = \{1, -1, -3\} \\
\Omega_4 &= \{TTHH, TTHT, TTTH, TTTT\} = \{0, -2, -4\} \\
&\vdots
\end{aligned}$$

Now consider the setup where you start with 80 dollars and bet the maximum amount of money you can per turn in order to reach 100 dollars as fast as possible. (Recall this was the optimal strategy – it maximized the probability of reaching 100 dollars before going bankrupt)



For this case we have

$$
\begin{aligned}
\Omega_1 &= \{20, -20\} \\
\Omega_2 &= \{20, -60\} \\
\Omega_3 &= \{-40, -80\} \\
\Omega_4 &= \{0, -80\}
\end{aligned}
$$

and it repeats starting from the fifth timestep:

$$\Omega_5 = \{20, -20\}$$

We will calculate the expected time to stop the game, whether we win the game or lose. Let t_i denote the time it takes to finish the game, starting with i dollars. Then we have the following set of equations:

$$
\begin{aligned}
t_{80} &= (p)1 + (1-p)t_{60} \\
t_{60} &= (p)1 + (1-p)t_{20} \\
t_{20} &= (p)t_{40} + (1-p)1 \\
t_40 &= (p)t_{80} + (1-p)1
\end{aligned}
$$

and our overall desired quantity is t_{80}.

When solved, we get:

$$
\begin{aligned}
\left[1 - p^2(1-p)^2\right] t_{80} &= p + p(1-p) + (1-p)^3 + p(1-p)^3 \\
\implies t_{80} &= \left[p + p(1-p) + (1-p)^3 + p(1-p)^3\right] / \left[1 - p^2(1-p)^2\right]
\end{aligned}
$$

Now calculate the probability of reaching 100 before 0 using such a betting strategy, starting from 80 dollars.

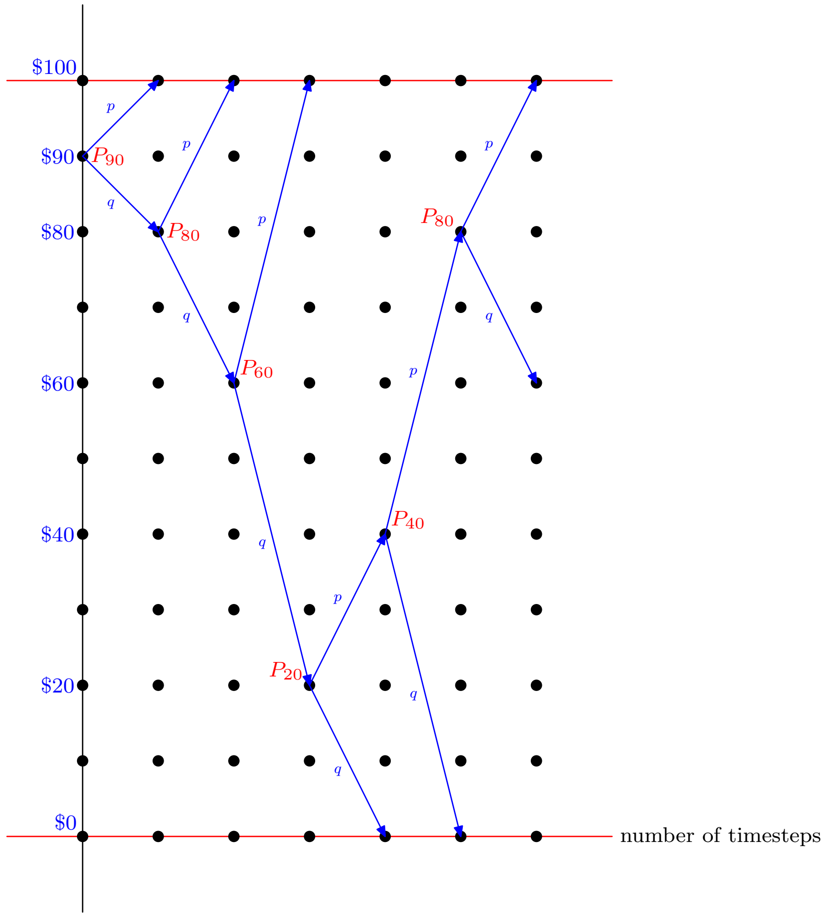
Enumerate every possible case:
We get the expression:

$$
\begin{aligned}
P(t_{80}) &= [p + p(1-p)] + p^3(1-p)\,[p + p(1-p)] + (p^3(1-p))^2\,[p + p(1-p)] + \cdots \\
&= [p + p(1-p)]\,\{1 + p^3(1-p) + (p^3(1-p))^2 + \cdots\} \\
&= [p + p(1-p)] * \left[1/(1 - p^3(1-p))\right] \\
&= \left[2p - p^2\right] * \left[1/(1 - p^3(1-p))\right]
\end{aligned}
$$

If we start from 90 dollars, then we must account for the extra probability p of winning directly and the probability $1-p$ of getting to 80 dollars, from which we can use our previously calculated probability. Overall expression is:

$$p + (1-p) * \left[2p - p^2\right] * \left[1/(1 - p^3(1-p))\right]$$

Now consider the setup where you start with 90 dollars and bet the maximum amount of money you can per turn in order to reach 100 dollars as fast as possible. (Recall this was the optimal strategy – it maximized the probability of reaching 100 dollars before going bankrupt). Recall the diagram was as follows:



In this case, we have

$$\Omega_1 = \{H, T\} = \{10, -10\}$$

where the payoffs change because of the amount of money we are betting per turn. Similarly:

$$\begin{aligned}
\Omega_2 &= \{TH, TT\} = \{-10+20, -10-20\} = \{10, -30\} \\
\Omega_3 &= \{TTH, TTT\} = \{-10, -70\} \\
\Omega_4 &= \{TTTH, TTTT\} = \{-50, -90\} \\
\Omega_5 &= \{TTTHT, TTTHH\} = \{-90, -10\} \\
\Omega_6 &= \{TTTHHH, TTTHHT\} = \{0, -30\}
\end{aligned}$$

and recall that the cycle after the sixth timestep repeats, starting from 80. We can take advantage of this symmetry in our calculations.

Coupon Collecting Problem:

start each problem with numerical values

Suppose each box of a certain brand of cereal contains a single prize coupon. There are a total of 3 distinct types of coupons, and each coupon in a box of cereal has probability $\frac{1}{3}$ of being any type. A complete set of coupons is one where there is at least one of each type of coupon in the set.

Let T be the random variable that describes the number of boxes of cereal that need to be bought in order to obtain a complete set of coupons. Determine the probability that we need exactly 5 coupons to obtain a complete set, i.e, $P(T=5)$.

In order for exactly 5 coupons to form a complete set, the fifth coupon must be the only one of its type. There are three different choices of the type of the last coupon.

Furthermore, in the four spots that remain, each spot can be occupied by a coupon that is either of the two remaining types. Hence, there are 2^4 remaining combinations. In order to get a complete set, there must be at least one of each type among the four slots. There are only two cases where this does not happen: either all four coupons are of one type or all four coupons are of the other type. Hence, we have $2^4 - 2$ possible complete set outcomes for a fixed third type in the fifth slot.

Overall, there are $3 * 2 * (2^3 - 1)$ possible complete sets within five coupons slots. There are a total of 3^5 coupon combinations for the five slots. Hence, the desired probability is $\frac{6(2^3-1)}{3^5}$.

Now let us generalize to a total of N distinct types of coupons, where each coupon in a box of cereal has probability $\frac{1}{N}$ of being any type. We'd like to determine the probability of needing exactly n coupons to obtain a complete set, i.e, $P(T=n)$.

Again, we will first do the easier computation of the probability that it takes more than n coupons in order to get a complete set, i.e, $P(T>n)$. Then:

$$P(T=n) = P(T>n-1) - P(T>n)$$

Define A_j to be the event that no coupon of type j is contained among the first n coupons. Then the event $\{T>n\}$ occurs when a Type 1 or Type 2 or $\cdots$ Type N coupon is not among the first n coupons seen. This event is denoted by

$$\bigcup_{i=1}^{N} A_j$$

and so

$$\begin{aligned}
P(T>n) &= P\left(\cup_{i=1}^{N} A_j\right) \\
&= \sum_j P\left(A_j\right) - \sum\sum_{j_1<j_2} P\left(A_{j_1}A_{j_2}\right) + \sum\sum\sum_{j_1<j_2<j_3} P\left(A_{j_1}A_{j_2}A_{j_3}\right) \\
&\quad - \cdots + \sum\cdots\sum_{j_1<\cdots<j_k} (-1)^{k+1} P\left(A_{j_1}\cdots A_{j_k}\right) + \cdots + (-1)^{N+1} P\left(Aj_1 \cdots A_{j_N}\right)
\end{aligned}$$

and note that for $n>1$, $P\left(Aj_1 \cdots A_{j_N}\right) = 0$ because the first coupon must have a type.

To compute $P\left(A_j\right)$, note that each of n coupons collected so far has a probability of $\frac{N-1}{N}$ of not being a Type j coupon. Hence,

$$P\left(A_j\right) = \left(\frac{N-1}{N}\right)^n$$

Similarly, each coupon has a probability of $\frac{N-2}{N}$ of not being Type j_1 or Type j_2. Hence,

$$P\left(A_{j_1}A_{j_2}\right) = \left(\frac{N-2}{N}\right)^n$$

and extending this result:

$$P\left(A_{j_1}\cdots A_{j_k}\right)=\left(\frac{N-k}{N}\right)^n$$

Plugging these into the overall expression above:

$$\begin{aligned}
P(T>n) &= \sum_j \left(\frac{N-1}{N}\right)^n - \sum\sum_{j_1<j_2}\left(\frac{N-2}{N}\right)^n + \sum\sum\sum_{j_1<j_2<j_3}\left(\frac{N-3}{N}\right)^n + \cdots \\
&= N\left(\frac{N-1}{N}\right)^n - N(N-1)\left(\frac{N-2}{N}\right)^n + \cdots N(N-1)\cdots(N-k+1)(-1)^{k+1}\left(\frac{N-k}{N}\right)^n
\end{aligned}$$

and simplifying:

$$P(T>n)=\sum_{k=1}^{N-1}\frac{N!}{(N-k)!}(-1)^{k+1}\left(\frac{N-k}{N}\right)^n$$

We can then plug this form of expression into our original desired probability:

$$\begin{aligned}
P(T=n) &= P(T>n-1)-P(T>n) \\
&= \sum_{k=1}^{N-1}\frac{N!}{(N-k)!}(-1)^{k+1}\left(\frac{N-k}{N}\right)^{n-1} - \sum_{k=1}^{N-1}\frac{N!}{(N-k)!}(-1)^{k+1}\left(\frac{N-k}{N}\right)^n
\end{aligned}$$

Now determine the number of distinct types of coupons in the first n coupons collected. We will denote this random variable D_n, so that what we desire to compute is the expression $P(D_n=k)$.

First, choose k distinct types. Then we can determine the probability that this set of k types is the set of distinct types in the first n coupons collected. In order for this to be the case, we must satisfy two conditions:

1. A: each coupon is one of the k types
2. B: there is at least one coupon of each type

Then we can express

$$P(D_n=k)=\binom{N}{k}P(AB)$$

since there are $\binom{N}{k}$ different groups of types of size k, each have the same probability $P(AB)$ of being the set that represents the n coupons.

Since each coupon selected has a probability $\frac{k}{N}$ of being one of the k types, we have

$$P(A)=\left(\frac{k}{N}\right)^n$$

Furthermore, given that a coupon is restricted to be one of k types, it is equally likely that this coupon is any one of these k types. Hence, $B|A$ denotes the event that there is at least one coupon of each of k types in a set of n coupons, where each coupon is equally likely to be any of the k types. But this is exactly the event that $\{T \leq n\}$, using the notation of T as previous. So:

$$\begin{aligned} P(B|A) &= 1 - P(T > n) \\ &= 1 - \sum_{i=1}^{k-1} \binom{k}{i} \left(\frac{k-i}{k}\right)^n (-1)^{i+1} \end{aligned}$$

Overall,

$$\begin{aligned} P(D_n = k) &= \binom{N}{k} P(A) P(B|A) \\ &= \binom{N}{k} \left(\frac{k}{N}\right)^n \left\{1 - \sum_{i=1}^{k-1} \binom{k}{i} \left(\frac{k-i}{k}\right)^n (-1)^{i+1}\right\} \end{aligned}$$

Determine the expected value of T. Recall T is the number of coupons to be collected until a complete set is obtained.

Let X be the number of different coupons in the set of n coupons and indicators

$$X_i = \begin{cases} 1 \text{ if } i\text{th type is in the set} \\ 0 \text{ else} \end{cases}$$

for $i = 1, \cdots, N$. Then we have

$$X = X_1 + X_2 + \cdots + X_N$$

Each of the X_i are iid, and

$$\begin{aligned} E(X_i) &= P(X_i = 1) \\ &= 1 - P(X_i = 0) \\ &= 1 - \left(\frac{N-1}{N}\right)^n \end{aligned}$$

Hence,

$$\begin{aligned} E(X) &= \sum_{i=1}^{N} \left[1 - \left(\frac{N-1}{N}\right)^n\right] \\ &= N \left\{1 - \left(\frac{N-1}{N}\right)^n\right\} \end{aligned}$$

Now determine the expected value of D_n. Recall D_n is the number of different types of coupon in the first n coupons collected.
Let Y denote the number of coupons that one must amass in order to get a complete set of N coupons. Further, let Y_i $(i = 0, 1, \cdots N-1)$ be the number of extra coupons that must be attained after i types have been collected such that another distinct type is obtained.

Then we have $Y = Y_0 + \cdots + Y_{N-1}$.

Each Y_i is independent from each other and geometric with parameter $\frac{N-i}{N}$, since after i types have been seen, there is a probability of $\frac{N-i}{N}$ of seeing a type you've never encountered before. Thus,

$$\begin{aligned} E(Y) &= \sum i = 1^N \frac{N}{N-i} \\ &= 1 + \frac{N}{N-1} + \frac{N}{N-2} + \cdots + N \end{aligned}$$

Additionally, the variance is easy to compute:

$$\begin{aligned} Var(Y) &= Var(Y_0) + Var(Y_1) + \cdots + Var(Y_{N-1}) \\ &= 0 + \frac{\frac{1}{N}}{\left(\frac{N-1}{N}\right)^2} + \frac{\frac{2}{N}}{\left(\frac{N-2}{N}\right)^2} + \cdots + \frac{\frac{N-1}{N}}{\left(\frac{1}{N}\right)^2} \\ &= \frac{N}{(N-1)^2} + \frac{2N}{(N-2)^2} + \cdots + (N-1)N \\ &= \sum_{i=1}^{N-1} \frac{iN}{(N-i)^2} \end{aligned}$$

Now suppose the type of a given coupon is not equally probable, i.e, a coupon is Type i with probability p_i and $\sum_i p_i = 1$. In this case, what is the expected number of coupons to get a complete set?

Use similar notation as previous. We want to determine $E(Y)$ and define Y_i to be the number of coupons needed to obtain a Type i coupon for the first time. Then Y_i is a geometric random variable with parameter p_i and

$$Y = \max_{i=1,\cdots,n} Y_i$$

We can use the following formula:

$$E(Y) = E\left(\sum_{i=1}^{N} Y_i\right) - E\left(\sum_{i<j} \min(Y_i, Y_j)\right) + \cdots + (-1)^{N+1} E\left(\min(Y_1, \cdots, Y_N)\right)$$

Note that $\min(Y_i, Y_j)$ is the number of coupons needed to obtain either a Type i or a Type j. This random variable is geometric with parameter $p_i + p_j$. Similarly, $\min(Y_i, Y_j, Y_k)$ is geometric with parameter $p_i + p_j + p_k$ and we can extend this pattern as the number of variables increases.

Hence

$$\begin{aligned} E(Y) &= E\left(\sum_{i=1}^{N} Y_i\right) - E\left(\sum_{i<j} \min(Y_i, Y_j)\right) + \cdots + (-1)^{N+1} E\left(\min(Y_1, \cdots, Y_N)\right) \\ &= \sum_{i=1}^{N} \frac{1}{p_i} - \sum_{i<j} \frac{1}{p_i + p_j} + \cdots + (-1)^{N+1} * \frac{1}{p_1 + \cdots + p_N} \end{aligned}$$

In this same setting, determine the expected value and variance of the number of different types of coupons that appear among the first n collected.

Let Y be the number of different types collected. For sake of ease, we will instead work with the number of uncollected types, denoted X. Then $X = N - Y$.

Define the indicator random variable

$$X_i = \begin{cases} 1 \text{ if there are no Type } i \text{ coupons in the collection} \\ 0 \text{ else} \end{cases}$$

Then

$$\begin{aligned} X &= \sum_{i=1}^{N} X_i \\ \implies E(X) &= \sum_{i=1}^{N} E(X_i) \end{aligned}$$

Note that

$$E(X_i) = P(X_i = 1) = (1 - p_i)^n$$

since the probability that a coupon is not of Type i is $1 - p_i$.

Hence,

$$\begin{aligned} E(X) &= \sum_{i=1}^{N} E(X_i) \\ &= \sum_{i=1}^{N} (1 - p_i)^n \end{aligned}$$

and

$$E(Y) = N - \sum_{i=1}^{N} (1 - p_i)^n$$

To calculate the variance, recall the formula

$$Var(Y) = E(Y^2) - E(Y)^2$$

We have that

$$\begin{aligned} E(Y^2) &= E(N^2 - 2NX + X^2) \\ &= N^2 - 2NE(X) + E(X^2) \\ &= N^2 - 2N \sum_{i=1}^{N} (1 - p_i)^n + E(X^2) \end{aligned}$$

Note that

$$
\begin{aligned}
E(X^2) &= \left(\sum_{i=1}^{N} E(A_i)\right)^2 \\
&= \sum_{i=1}^{N} E({A_i}^2) + 2\sum_{i<j} E(A_i A_j) \\
&= \sum_{i=1}^{N} E(A_i) + 2\sum_{i<j}(1-p_i-p_j)^n \\
&= \sum_{i=1}^{N}(1-p_i)^n + 2\sum_{i<j}(1-p_i-p_j)^n
\end{aligned}
$$

where $E({A_i}^2) = E(A_i)$ because A_i is an indicator and A_iA_j denotes the event that all n coupons are neither Type i nor Type j, which occurs with probability $(1-p_i-p_j)^n$.

Substituting back:

$$
\begin{aligned}
E(Y^2) &= N^2 - 2N\sum_{i=1}^{N}(1-p_i)^n + \sum_{i=1}^{N}(1-p_i)^n + 2\sum_{i<j}(1-p_i-p_j)^n \\
&= N^2 - (2N-1)\sum_{i=1}^{N}(1-p_i)^n + 2\sum_{i<j}(1-p_i-p_j)^n
\end{aligned}
$$

and we can use the above formula for variance to achieve our desired result.

Now go back to the scenario where the probability of a coupon being a certain Type is equal and let X denote the number of Types for which exactly one coupon of that type is collected. Determine $E(X), Var(X)$.

Let T_i be the ith new type of coupon to be collected in the sequence of n coupons. Define the indicator

$$
X_i = \begin{cases} 1 \text{ if type } T_i \text{ has exactly one coupon} \end{cases}
$$

so that

$$
\begin{aligned}
X &= \sum_{i=1}^{N} X_i \\
\implies E(X) &= \sum_{i=1}^{N} E(X_i)
\end{aligned}
$$

At the moment the first Type T_i coupon is collected, there remain $N-i$ types that still need to be obtained in order to form a complete set.

Appendix:

Here we summarize all calculations and results referred to in the Gambler's Ruin Problem above. All computations were done using MATLAB.

1. Set up a matrix equation of the form $Ax = b$ and let x be the vector of probabilities p_i with i from 1 to 19. Then we have:

```
% probability for 9/10-10/19 two steps
A4 = zeros(19);
for i = 2:18
    A4(i,i - 1) = -10/19;
    A4(i,i) = 1;
    A4(i, i+1) = -9/19;
end
A4(1,1) = 1;
A4(1,2) = -9/19;
A4(19,18) = -10/19;
A4(19,19) = 1;

b4 = zeros(1,19)';
b4(19,1) = 9/19;
x4 = A4\b4;
```

We obtain the following x:

```
x4 =

    0.0154
    0.0325
    0.0515
    0.0725
    0.0960
    0.1220
    0.1510
    0.1831
    0.2188
    0.2585
    0.3026
    0.3516
    0.4061
    0.4666
    0.5338
    0.6085
    0.6915
    0.7837
    0.8862
```

2. Set up a matrix equation of the form $Ax = b$ and let x be the vector of mean times m_i with i from 1 to 19. Then we have:

```
% mean time to absorption for 9/19-10/19 two steps
A2 = zeros(19);
for i = 2:18
    A2(i,i - 1) = -10/19;
    A2(i,i) = 1;
    A2(i, i+1) = -9/19;
end
A2(1,1) = 1;
A2(1,2) = -9/19;
A2(19,18) = -10/19;
A2(19,19) = 1;

b2 = ones(1,19)';
x2 = A2\b2;
```

We obtain the following x:

```
x2 =

   13.1563
   25.6633
   37.4489
   48.4328
   58.5261
   67.6298
   75.6338
   82.4161
   87.8409
   91.7573
   93.9978
   94.3760
   92.6852
   88.6955
   82.1513
   72.7688
   60.2328
   44.1927
   24.2593
```

3. Set up a matrix equation of the form $Ax = b$ and let x be the vector of probabilities p_i with i from 1 to 49.

```
% probability for 9/10-10/19 five steps
A5 = zeros(49);
for i = 2:48
    A5(i,i - 1) = -10/19;
    A5(i,i) = 1;
    A5(i, i+1) = -9/19;
end
A5(1,1) = 1;
A5(1,2) = -9/19;
A5(49,48) = -10/19;
A5(49,49) = 1;

b5 = zeros(1,49)';
b5(49,1) = 9/19;
x5 = A5\b5;
```

We obtain the following x:

```
x5 =

    0.0006
    0.0012
    0.0019
    0.0027
    0.0036
    0.0046
    0.0057
    0.0069
    0.0082
    0.0097
    0.0113
    0.0132
    0.0152
    0.0175
    0.0200
    0.0228
    0.0259
    0.0293
    0.0332
    0.0374
    0.0422
    0.0474
    0.0533
    0.0598
```

```
0.0670
0.0750
0.0839
0.0938
0.1048
0.1170
0.1306
0.1457
0.1625
0.1811
0.2018
0.2248
0.2503
0.2787
0.3103
0.3453
0.3842
0.4275
0.4756
0.5290
0.5884
0.6543
0.7276
0.8090
0.8995
```

4. Here, x is the vector of mean times m_i with i from 1 to 49.

```
% mean time to absorption for 9/19-10/19 five steps
A6 = zeros(49);
for i = 2:48
    A6(i,i - 1) = -10/19;
    A6(i,i) = 1;
    A6(i, i+1) = -9/19;
end
A6(1,1) = 1;
A6(1,2) = -9/19;
A6(49,48) = -10/19;
A6(49,49) = 1;

b6 = ones(1,49)';
x6 = A6\b6;
```

We obtain the following x:

```
x6 =

   18.4532
   36.8456
   55.1705
   73.4204
   91.5869
  109.6609
  127.6319
  145.4886
  163.2183
  180.8069
  198.2386
  215.4960
  232.5599
  249.4086
  266.0183
  282.3624
  298.4114
  314.1325
  329.4892
  344.4412
  358.9434
  372.9459
  386.3930
  399.2232
  411.3678
  422.7508
  433.2874
  442.8836
  451.4349
  458.8253
  464.9257
  469.5929
  472.6675
  473.9726
  473.3116
  470.4661
  465.1932
  457.2234
  446.2570
  431.9609
  413.9653
  391.8591
  365.1855
  333.4370
  296.0499
  252.3975
  201.7837
  143.4351
   76.4921
```

5. In matrix form, this is:

```
A = [1 -9/19 0 0 0 0 0 0 0 0;
     0 1 0 -9/19 0 0 0 0 0 0;
     -10/19 0 1 0 0 0 0 0 0 0;
     0 0 -10/19 1 0 0 0 0 0 0;
     0 0 0 -10/19 1 0 0 0 0 0;
     0 0 0 0 0 1 -9/19 0 0 0;
     0 0 0 0 0 0 1 0 -9/19 0;
     0 0 0 0 0 -10/19 0 1 0 0;
     0 0 0 0 0 0 0 -10/19 1 0;
     0 0 0 0 0 0 0 0 -10/19 1;];

b = [10/19 10/19 0 0 0 0 0 9/19 9/19 9/19]';
x = A\b
```

and we obtain the following vector of probabilities. Note that the results are the same as previously computed.

```
x =

    0.8270
    0.6348
    0.4353
    0.2291
    0.1206
    0.1730
    0.3652
    0.5647
    0.7709
    0.8794
```

In the case where $p = \frac{1}{2}$, all is the same as above, with all transition probabilities replaced with $\frac{1}{2}$.

```
A = [1 -1/2 0 0 0 0 0 0 0 0;
     0 1 0 -1/2 0 0 0 0 0 0;
     -1/2 0 1 0 0 0 0 0 0 0;
     0 0 -1/2 1 0 0 0 0 0 0;
     0 0 0 -1/2 1 0 0 0 0 0;
     0 0 0 0 0 1 -1/2 0 0 0;
     0 0 0 0 0 0 1 0 -1/2 0;
     0 0 0 0 0 -1/2 0 1 0 0;
     0 0 0 0 0 0 0 -1/2 1 0;
     0 0 0 0 0 0 0 0 -1/2 1;];

b = [1/2 1/2 0 0 0 0 0 1/2 1/2 1/2]';
x = A\b
```

and we obtain the following vector of probabilities. Note that the results are the same as previously computed.

```
x =

    0.8000
    0.6000
    0.4000
    0.2000
    0.1000
    0.2000
    0.4000
    0.6000
    0.8000
    0.9000
```

We will use this result later on in this script.

6. In matrix form, this is:

```
A = [.9 -.8*.5 0 0 0 0 0 0 0 0;
     0 .8 -.6*.5 0 0 0 0 0 0 0;
     0 0 .6 0 -.2*.5 0 0 0 0 0;
     0 -.8*.5 0 .4 0 0 0 0 0 0;
     0 0 0 -.5*.4 .2 0 0 0 0 0;
     0 0 0 0 0 .1 -.5*.2 0 0 0;
     0 0 0 0 0 0 .2 -.5*.4 0 0;
     0 0 0 0 0 0 0 .4 0 -.5*.8;
     0 0 0 0 0 0 -.5*.2 0 .6 0;
     0 0 0 0 0 0 0 0 -.5*.6 .8]

b = [.9 .8 .6 .4 .2 .1 .2 .4 .6 .8]';
x = A\b
```

and we obtain the following vector of expected values.

```
x =

    1.7111
    1.6000
    1.6000
    2.6000
    3.6000
    4.6000
    3.6000
    2.6000
    1.6000
    1.6000
```

Chapter 2

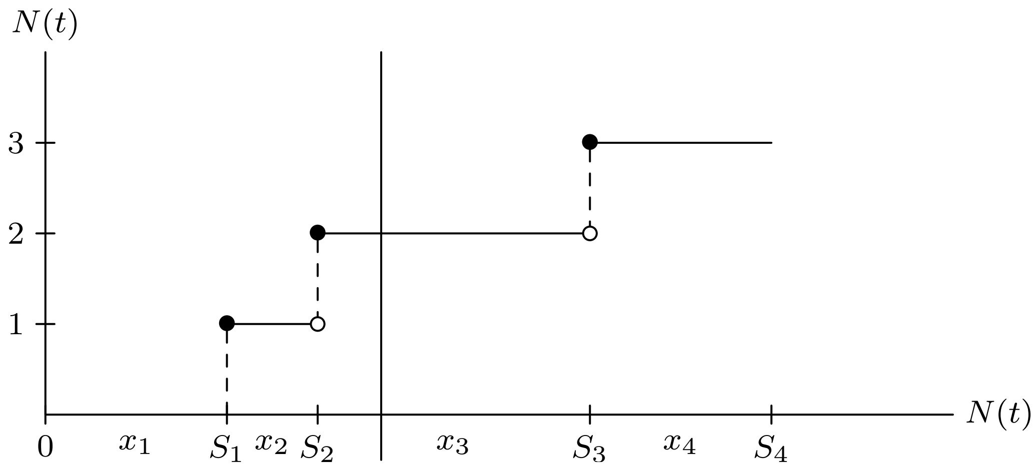
This chapter is revolves around problems and concepts that are slightly outside of the scope of EE 126, particularly the Patterns section, which is extended much further here than what was done in the course.

Renewal Processes Problems

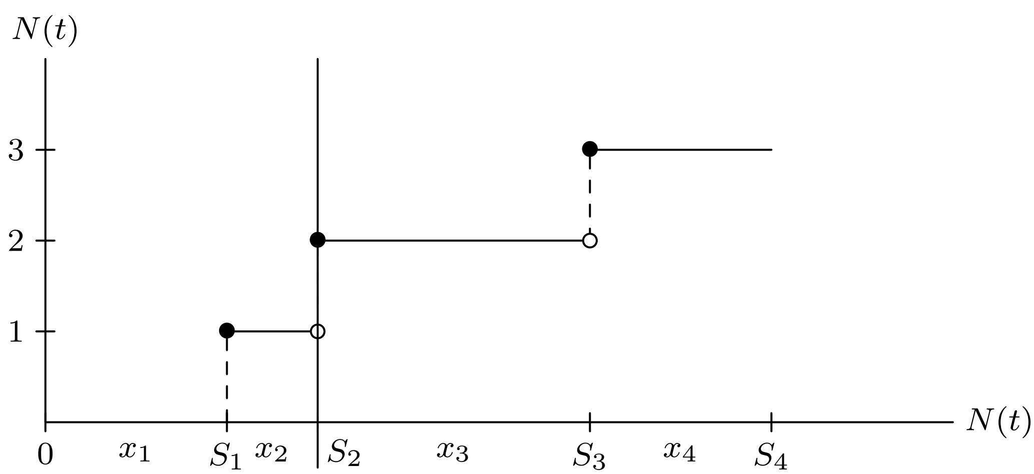
1. Is it true that

   (a) $N(t) < n$ iff $S_n > t$?
   Solution: True. Recall that $N(t) \geq n$ iff $S_n \leq t$ is true. Hence, the contrapositive must also be true.

   (b) $N(t) \leq n$ iff $S_n \geq t$?
   Solution: False. Consider the graph below. Suppose $t > S_2$, then $N(t) \leq 2$.



   (c) $N(t) > n$ iff $S_n < t$?
   Solution: False. Consider the graph below. If $t = S_2$, we have that $N(t) = 2$.



2. **Wald's Equation:** Let $X_1, X_2, \cdots$ be a sequence of independent random variables. The nonnegative integer valued random variable N is said to be a *stopping time* for the sequence if the event $\{N = n\}$ is independent of $X_{n+1}, X_{n+2}, \cdots$.

(a) Let X_i be Bernoulli random variables with parameter p, for all $i \geq 1$. Define the following:

$$N_1 = \min\{n : X_1 + \cdots X_5 = 5\}$$

$$N_2 = \begin{cases} 3, & \text{if } X_1 = 0 \\ 5, & \text{if } X_1 = 1 \end{cases}$$

$$N_3 = \begin{cases} 3, & \text{if } X_4 = 0 \\ 2, & \text{if } X_4 = 1 \end{cases}$$

Which of the N_i are stopping times for the sequence $X_1, \cdots$?

Solution: Note that N_1 essentially denotes the first time we obtain five successes. Then it is a stopping time. In both cases of N_2, there is independence of the events X_3 and X_5, as well as from all the events with increasing index. By definition, N_2 is also a stopping time. For N_3, both cases are dependent on X_4. Hence, it is not a stopping time.

(b) Wald's Equation states that $X_1, X_2, \cdots$ are independent and identically distributed and have a finite mean $E(X)$, and if N is a stopping time for this sequence having a finite mean, then

$$E\left[\sum_{i=1}^{N} X_i\right] = E\left[N\right] E\left[X\right]$$

To prove Wald's Equation, define indicator variables $I_i, i \geq 1$ by

$$I_i = \begin{cases} 1, & \text{if } i \leq N \\ 0, & \text{if } i > N \end{cases}$$

Then we can write

$$\sum_{i=1}^{N} X_i = \sum_{i=1}^{\infty} X_i I_i$$

which implies that

$$E\left[\sum_{i=1}^{N} X_i\right] = E\left[\sum_{i=1}^{\infty} X_i I_i\right] = \sum_{i=1}^{\infty} E\left[X_i I_i\right]$$

Since X_i and I_i are independent (the value of I_i is completely determined by $X_1, \cdots, X_{i-1}$ because I_i is 0 or 1 depending on whether or not we have stopped after observing up to X_{i-1}), we have that

$$\sum_{i=1}^{\infty} E\left[X_i I_i\right] = \sum_{i=1}^{\infty} E\left[X_i\right] E\left[I_i\right] = \sum_{i=1}^{\infty} E\left[X\right] E\left[I_i\right]$$

And from here, we obtain Wald's Equation.

$$\sum_{i=1}^{\infty} E\left[X\right] E\left[I_i\right] = E\left[X\right] \sum_{i=1}^{\infty} E\left[I_i\right] = E\left[X\right] E\left[N\right]$$

3. Recall the Elementary Renewal Theorem:

$$\frac{E[N(t)]}{t} \to \frac{1}{\mu} \text{ as } t \to \infty$$

Wald's Equation can be used as the basis pf a proof of this Elementary Renewal Theorem.

Let $X_1, X_2, \cdots$ denote the interarrival times of a renewal process and let $N(t)$ be the number of renewals by time t.

(a) Show that whereas $N(t)$ is not a stopping time, $N(t)+1$ is.
Solution: Note that $N(t) = n$ iff $X_1 + \cdots + X_n \leq t$ and $X_1 + \cdots + X_{n+1} > t$. This means that the event $\{N(t) = n\}$ is not independent of X_{n+1}. However, $\{N(t)+1 = n\} = \{N(t) = n-1\}$ iff $X_1 + \cdots + X_{n-1} \leq t$ and $X_1 + \cdots + X_n > t$ is independent of X_{n+1}.

(b) Applying Wald's Equation gives

$$\begin{aligned} E\left[\sum_{i=1}^{N(t)+1} X_i\right] &= E\left[X_i\right] E\left[N(t)+1\right] \\ &= \mu(m(t)+1) \end{aligned}$$

Suppose that the X_i are bounded random variables. That is, suppose there is a constant M such that $P(X_i < M) = 1$. Then we have that $t < \sum_{i=1}^{N(t)+1} X_i < t+M$. This is because $\sum_{i=1}^{N(t)+1} X_i$ is the time of the first renewal after t. The inequality come from the interpretation that there must be at least one renewal in the interval between t and $t+m$.

Taking expectation across the inequality, we get

$$\begin{aligned} t = E\left[t\right] < \mu(m(t)+1) < E\left[t+m\right] = t+m \\ \Longrightarrow t - \mu < \mu m(t) < t + M - \mu \\ \Longrightarrow \frac{1}{\mu} - \frac{1}{t} < \frac{m(t)}{t} < \frac{1}{\mu} + \frac{M-\mu}{\mu t} \end{aligned}$$

Note that as $t \to \infty$, $\frac{m(t)}{t} \to \frac{1}{\mu}$ by sandwiching.

4. Wald's Equation can also be proven by using Renewal Reward Processes.

(a) Let $N_1 = N$. Note that the sequence of random variables $X_{N_1+1}, X_{N_1+2}, \cdots$ is independent of $X_1, \cdots X_{N_1}$, and has the same distribution as the original sequence of X_i's, because of the definition of a stopping time.

Now let N_2 be the stopping time for the sequence $X_{N_1+1}, X_{N_1+2}, \cdots$, N_3 be the stopping time for the sequence $X_{N_1 N_2+1}, X_{N_1 N_2+2}, \cdots$, and so forth. Consider a reward process for which X_i is the reward earned during a period i. Then the process is a renewal reward process with cycle lengths $N_1, N_2, \cdots$.

By the Renewal Reward Theorem, the average reward per unit time is $\frac{E[X_1 + \cdots X_N]}{E(N)}$.

By the Strong Law of Large Numbers, we have

$$\lim_{n\to\infty} \sum_{i=1}^{n} \frac{X_i}{n} \to E(X)$$

Hence, $E(X) = \frac{E[X_1 + \cdots X_N]}{E(N)}$, and we have Wald's Equation.

5. Consider a miner trapped in a room that has three doors. One door leads him to freedom after a two-day walk. Another door leads him back to the room after a four-day walk, and the last door leads him back to the room after a six-day-walk. Suppose at all times, he is equally likely to choose any of the three doors. Let T denote the time it takes for the miner to become free.

   Define X_i to be the amount of time he travels after his ith door pick.

$$X_i = \begin{cases} 2, \text{w.p } \frac{1}{3} \\ 4, \text{w.p } \frac{1}{3} \\ 6, \text{w.p } \frac{1}{3} \end{cases}$$

   Further define the stopping time N such that $T = \sum_{i=1}^{N} X_i$, i.e, the time it takes for him to get free.

   (a) We will first compute T the "normal" way:
   Set up the following equation:

$$E(T) = \tfrac{1}{2}(2 + (4 + E(T)) + (6 + E(T)))$$
$$\implies E(T) = 12$$

   (b) Now use Wald's Equation and see if we get the same answer:

$$\begin{aligned} E\left[\sum_{i=1}^{N} X_i\right] &= E(X_i)E(N) \\ &= \frac{1}{3}(2+4+6) * 3 \\ &= 12 \end{aligned}$$

   (c) Compute $E\left[\sum_{i=1}^{N} X_i | N = n\right]$.

   Solution: Given $N = n$, it must be the case that the first $n-1$ trials were failures. And failures only come in the form of a four-day journey or a six-day journey, with equal probability. Hence:

$$E\left[\sum_{i=1}^{N} X_i | N = n\right] = (n-1)(\tfrac{1}{2}(4+6)) + 2 = 5n - 3$$

   and the 2 at the end comes from the two-day journey that leads to freedom on the nth trial.

   Note on the other hand:

$$E\left[\sum_{i=1}^{n} X_i\right] = n * \tfrac{1}{3}(2+4+6) = 4n$$

. Thus,

$$E\left[\sum_{i=1}^{n} X_i\right] \neq E\left[\sum_{i=1}^{N} X_i | N = n\right]$$

(d) Now $E(T) = E\left[\sum_{i=1}^{N} X_i\right] = E(5N-3) = 15-3 = 12$. This concludes the third way of deriving $E(T)$.

6. A deck of 52 playing cards is shuffled and the cards are then turned face up one at a time. Define X_i to be the indicator variable for when the ith card turned over is an ace. Let N denote the number of cards that were turned over until all four aces appear.

   Note that the X_i are not independent because the total number of cards decreases with every card that is turned face-up, and the chance that a card is an ace varies. Hence, Wald's Equality cannot be applied.

   To check, we have that the left side of Wald's Equality is 4 (the only contribution to the sum is when a card is an ace, and we stop when we've flipped four of them) and the right side is $E(N) * \frac{1}{13}$ (the probability that a card is an ace is $\frac{1}{13}$). This means that $E(N) = 52$, i.e, we have to go through the whole pile of cards in order to get all four aces. This is clearly incorrect. Indeed, Wald's Equality cannot be applied in this case.

7. Consider the gambler's ruin problem where on each bet, the gambler either gains a dollar (+1) with probability p, or loses a dollar (−1) with probability $1-p$. Starting with i dollars, the gambler will continue playing until the total of his new winnings are either $N-i$ or $-i$. Let T be the number of bets made until the gambler stops. Use Wald's Equation, along with the known probability that the gambler's final winnings are $N-i$, to find $E(T)$.

8. A worker works sequentially on jobs. Each time a job is completed, a new one is begun. Each job, independently, takes a random amount of time having distribution F to complete. However, independently of this, shocks occur according to a Poisson process with rate λ. Whenever a shock occurs, the worker discontinues his current job to start a new one. In the long run, at what rate are jobs completed?

9. Consider a renewal process with mean interarrival time μ. Suppose each event of this process is independently "counted" with probability p. Let $N_C(t)$ denote the number of counted events by time t, $t > 0$.

   (a) Is $N_C(t), t \geq 0$ a renewal process?

   (b) What is $\lim_{t\to\infty} \frac{N_C(t)}{t}$?

10. Events occur according to a Poisson process with rate λ. Any event that occurs within a time d of the event that immediately preceded it is called a d-event. For example, if $d = 1$ and events occur at times $2, 2.8, 4, 6, 6.6, \cdots$, then the events at times 2.8 and 6.6 would be d-events.

    (a) At what rate do d-events occur?

(b) What proportion of all events are d-events?

11. Consider a train station to which customers arrive in accordance with a Poisson process having rate λ. A train is called whenever there are N customers waiting at the station, but it takes K units of time for the train to arrive at the station. When it arrives, it picks up all waiting customers. Assume that the train station incurs a cost at a rate of nc per unit time whenever there are n customers present. Find the long-run average cost.

12. A machine consists of two independent components, the ith of which functions for an exponential time with rate λ_i. The machine functions as long as at least one of these components function. When a machine fails, a new machine having both its components working is put into use. A cost K is incurred whenever a machine failure occurs; operating costs at rate c_i per unit time are incurred whenever the machine in use has i working components, for $i = 1, 2$. Find the long-run average cost per unit time.

13. Consider a single server queueing system in which customers arrive in accordance with a renewal process. Each customer brings in a random amount of work, chosen independently according to a distribution G. The server serves one customer at a time. However, the server processes work at rate i per unit time whenever there are i customers in the system. For instance, if a customer with workload 8 enters service when there are three other customers waiting in line, and no one else arrives, then that customer will spend 2 units of time in service. If another customer arrives after 1 unit of time, then that customer will spend a total of 1.8 units of time in service, provided no one else arrives. Let W_i denote the amount of time customer i spends in the system. Also, define $E(W)$ to be the average amount of time a customer spends in the system:

$$E(W) = \lim_{n\to\infty} \frac{W_1 + \cdots W_n}{n}$$

Patterns

1. Suppose we had a two-sided coin that lands heads with probability p. We toss it until we get a succession of k heads in a row. Find the expected number of tosses we must make.

   Solution: Use the regular definition of expectation:

$$E(T) = \sum_{i=1}^{k} iP(T = i)$$

   Condition on the time of the first tail and rewrite the above equation:

$$E(T) = \sum_{i=1}^{k} (1-p)p^{i-1}(i + E(T)) + kp^k$$

Solve for $E(T)$:

$$\begin{aligned}
E(T) &= (1-p)\sum_{i=1}^{k} ip^{i-1} + E(T)(1-p)\sum_{i=1}^{k} p^{i-1} + kp^k \\
&= (1-p)\sum_{i=1}^{k} ip^{i-1} + E(T)(1-p)\frac{1-p^k}{1-p} + kp^k \\
p^k E(T) &= (1-p)\sum_{i=1}^{k} ip^{i-1} + kp^k \\
E(T) &= \frac{1-p}{p^k}\sum_{i=1}^{k} ip^{i-1} + k \\
&= \frac{1}{p^k}\sum_{i=0}^{k-1} p^i = \frac{(1-p^k)}{p^k(1-p)}
\end{aligned}$$

Now extend the above example to the following problem:

A sequence of independent trials, each of which results in outcome number i with probability $P_i, i = 1, 2, \cdots n$ such that $\sum_{i=1}^{n} P_i = 1$, is observed until the same outcome occurs k times in a row. This outcome is then declared to be the winner of the game.

For instance, if $k = 2$ and the sequence of outcomes is $1, 2, 4, 3, 5, 2, 1, 3, 3$, then we stop after nine trials and declare outcome number 3 the winner.

What is the probability that i wins, $i = 1, \cdots, n$, and what is the expected number of trials?

2. We are tossing a coin that lands a success with probability p, and fails with probability $q = 1 - p$. Compute the probability of getting n consecutive successes before m consecutive failures.

Solution: Let A be the event that we get n consecutive successes before m consecutive failures. Let S denote the event that a success occurred in the first toss, and F denote the event that a failure occurred instead. By conditional probabilities, we get:

$$\begin{aligned}
P(A) &= P(A|S) * P(S) + P(A|F) * P(F) \\
&= P(A|S) * p + P(A|F) * q
\end{aligned}$$

Let us compute an expression for $P(A|S)$. From the first success, we need $n-1$ more success in order to get n consecutive successes. Then the probability of event A occurring is 1. On the other hand, if at least one failure occurred within the $n-1$ subsequent tosses, then we break the desired sequence, and must restart the whole process, starting with the first failure in there. Hence:

$$\begin{aligned}
P(A|S) &= p^{n-1} * 1 + (1 - p^{n-1}) * P(A|F) \\
&= p^{n-1} + (1 - p^{n-1})P(A|F)
\end{aligned}$$

We can similarly compute an expression for $P(A|F)$. From the first failure, we need $m-1$ more failures in order to get m consecutive failures. Then the probability of event A occurring is 0. On the other hand, if at least one success occurred within the $m-1$ subsequent tosses, then we break the string of failures, and must restart the whole process, starting with the first success in there. Hence:

$$\begin{aligned} P(A|F) &= q^{m-1} * 0 + (1-q^{m-1}) * P(A|S) \\ &= (1-q^{m-1})P(A|S) \end{aligned}$$

We have two equations and two unknowns, $P(A|S)$ and $P(A|F)$. Solving for them yields:

$$\begin{aligned} P(A|S) &= \frac{p^{n-1}}{p^{n-1}+q^{m-1}-p^{n-1}q^{m-1}} \\ P(A|F) &= \frac{(1-q^{m-1})p^{n-1}}{p^{n-1}+q^{m-1}-p^{n-1}q^{m-1}} \end{aligned}$$

Substituting into the expression for our original desired probability, we get:

$$\begin{aligned} P(A) &= P(A|S) * p + P(A|F) * q \\ &= \frac{q^{m-1}(1-p^n)}{p^{n-1}+q^{m-1}-p^{n-1}q^{m-1}} \end{aligned}$$

3. n men take off their hats and throw it into the air. Once the hats have landed on the ground, they pick one up at random. The event that a man picks up his own hat is called a "match". What is the probability of having:

   (a) no matches?
   Solution: Let A_i denote the event that each of i men have picked up a hat that does not belong to him. Then we desire to compute the probability of A_n.

   Let M_1 be the event that the "first" man has his own hat. Then the complement, $\overline{M}_1$ is the event that the first man does not have his own hat. Conditioning on these events, we get:

   $$P(A_n) = P(A_n|M_1)P(M_1) + P(A_n|\overline{M}_1)P(\overline{M}_1)$$

   Note that $P(A_n|M_1) = 0$ since we cannot have n unmatching man-hat pairs if the first man has his own hat. Hence, the first term of the sum cancels out. Also, we have that $P(\overline{M}_1) = \frac{n-1}{n}$.

   To calculate $P(A_n|\overline{M}_1)$, note that the event $A_n|\overline{M}_1$ implies that the first man holds one of the hats of the other $n-1$ men. Denote M_2 to be the event that the first man is holding the "second" man's hat.

   Divide this further into two mutually exclusive events: (1) there are no matches and the second man has the first man's hat, and (2) there are no matches and the second man does not have the first man's hat. Then we have:

   $$P(A_n|\overline{M}_1) = \frac{1}{n-1}P(A_{n-2}) + P(A_{n-1})$$

Substituting, we have:

$$\begin{aligned} P(A_n) &= \left(\frac{1}{n-1} P(A_{n-2}) + P(A_{n-1}) \right) \frac{n-1}{n} \\ &= \frac{1}{n} P(A_{n-1}) + \frac{n-1}{n} P(A_{n-1}) \end{aligned}$$

We can massage this into a more recursive expression:

$$P(A_n) - P(A_{n-1}) = -\frac{1}{n} \left[P(A_{n-1}) - P(A_{n-2}) \right]$$

We start with $P(A_1) = 0$ and $P(A_2) = \frac{1}{2}$. Using the above, we get $P(A_3) = -\frac{1}{3!} + \frac{1}{2!}$ and $P(A_4) = \frac{1}{4!} - \frac{1}{3!} + \frac{1}{2!}$.

Eventually, we reach our final answer:

$$P(A_n) = \frac{(-1)^n}{n!} + \cdots + \frac{1}{4!} - \frac{1}{3!} + \frac{1}{2!}$$

(b) k matches?
Solution: Any group of k men, among the n total must select their own hats. This occurs with probability:

$$\binom{n}{k} \frac{1}{n} * \frac{1}{n-1} * \cdots * \frac{1}{n-k+1}$$

Furthermore, the remaining $n-k$ men must not choose their own hats. This occurs with probability $P(A_{n-k})$.

Hence we have overall probability:

$$\begin{aligned} &P(A_{n-k}) * \binom{n}{k} \frac{1}{n} * \frac{1}{n-1} * \cdots * \frac{1}{n-k+1} \\ &= \frac{\frac{(-1)^{n-k}}{(n-k)!} + \cdots + \frac{1}{4!} - \frac{1}{3!} + \frac{1}{2!}}{k!} \end{aligned}$$

4. We are tossing a coin that lands heads with probability p, and tails with probability $q = 1-p$. Compute the expected number of coin tosses it takes to get the sequence HHT.

Solution: Let A be the event where we get the desired sequence. Condition on whether we get a heads or a tails on the first toss.

$$\begin{aligned} E(A) &= E(A|H)P(H) + E(A|T)P(T) \\ &= E(A|H)p + E(A|T)q \end{aligned}$$

Note that $E(A|T) = 1 + E(A)$ since we want a heads in the first toss in order to begin counting for our sequence, but we've already tossed the coin once, so we add a one.

Now $E(A|H)$ can be further conditioned on whether we get a heads or a tails on the second toss.

$$\begin{aligned} E(A|H) &= E(A|HH)P(H|H) + E(A|HT)P(T|H) \\ &= E(A|HH)p + E(A|HT)q \end{aligned}$$

By similar reasoning as above, we have that $E(A|HT) = 2 + E(A)$. Also, $E(A|HH) = 2 + \frac{1}{q}$, since now that we have the first two heads, all we have to wait is for the first tail to land, and the number of tosses required for such is a geometric random variable with parameter q.

Hence

$$\begin{aligned} E(A|H) &= (2 + \frac{1}{q})p + (2 + E(A))q \\ &= 2p + \frac{p}{q} + 2q + qE(A) \\ &= 2 + \frac{p}{q} + qE(A) \end{aligned}$$

Substituting this into our original expression for $E(A)$, we get:

$$\begin{aligned} E(A) &= (2 + \frac{p}{q} + qE(A))p + (1 + E(A))q \\ &= 2p + \frac{p^2}{q} + qpE(A) + q + qE(A) \\ (1 - qp - q)E(A) &= 2p^2 + \frac{p^2}{q} + 2pq + q \\ p^2E(A) &= 2p^2 + \frac{p^2}{q} + 2pq + q \end{aligned}$$

Now we get:

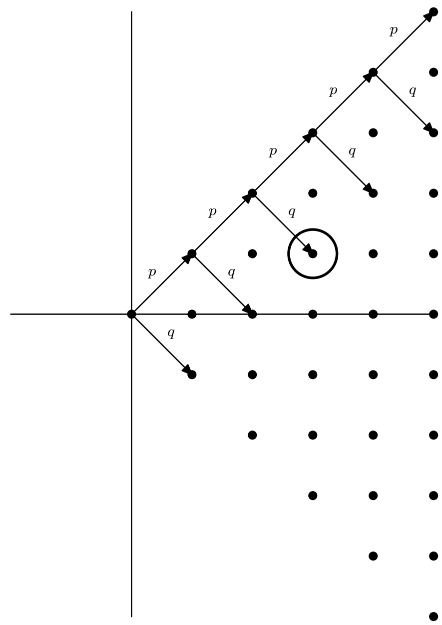
$$E(A) = \frac{2p^2q + p^2 + 2q^2p + q^2}{p^2q}$$

Note that the numerator of this expression can be simplified as follows:

$$\begin{aligned} &2p^2q + p^2 + 2q^2p + q^2 \\ &= 2p^2(1-p) + p^2 + 2(1-p)^2p + (1-p)^2 \\ &= 1 \end{aligned}$$

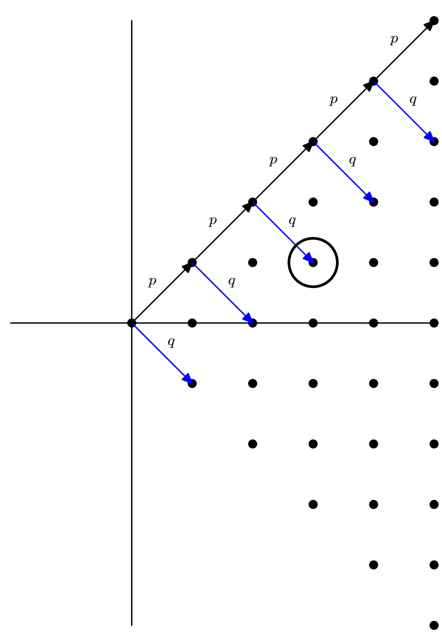
So our final answer is $\frac{1}{p^2q}$.

We will now consider a slightly different approach to solve the problem. We will be using the plot below for our computation:



The horizontal axis marks the number of tosses made. We go up one unit if our coin lands heads and down one unit if it lands tails. Hence, we have probabilities associated with every segment of the trajectory, as labeled above. If we were to obtain the sequence immediately since we start tossing the coin, the endpoint of the trajectory we desire would be as circled above.

We can compute the expected value of obtaining our desired sequence HHT from every point. First we consider the following trajectory:



For a more general desired sequence $\underbrace{H \cdots H}_{k} T$, the expected number of tosses required is $\frac{1}{p^k q}$.

Note: This is a simple example of computing the expected number of trials needed to observe a pattern with no repetition (meaning the first k consecutive outcomes and the last k consecutive outcomes of a pattern are not the same for all k). As we can easily calculate, the probability of getting HHT is p^2q, and so, the expected number of tosses needed to observe this sequence is $\frac{1}{p^2q}$. Similarly, if we wish to compute this quantity for the pattern $HHTHTHTT$, then it is just $\frac{1}{P(HHTHTHTT)} = \frac{1}{p^4q^4}$.

For cases where repetition does occur (for instance, HTH, where $k = 1$ and the first and last outcomes are both heads), we will consider via the following example:

5. Toss a fair die that has one red face, two blue faces, and three white faces. Find the expected number of rolls needed to observe the sequence of face colors: $RBWRB$.

Solution: First let us note that:

$$P(R) = \tfrac{1}{6}$$
$$P(B) = \tfrac{1}{3}$$
$$P(W) = \tfrac{1}{2}$$

Let N be the number of rolls of the die until we observe such a sequence of face colors. Note that $k = 2$ and we have the repeated sequence RB. So compute instead the expected value of M, the number of rolls needed to observe the sequence $RBWRBB$. Note that this pattern has no repetition, so we can apply our previous logic.

Then we have that $E(M) = 6^2 * 3^3 * 2$.

Let L be the additional number of rolls after the occurrence of the sequence $RBWRB$ until $RBWRBB$ occurs. Then we have that:

$$M = N + L \Longrightarrow E(M) = E(N) + E(L)$$

and we can compute the $E(L)$ as follows:

$$\begin{aligned} E(L) &= E(L|X = R)P(R) + E(L|X = B)P(B) + E(L|X = W)P(W) \\ &= \frac{1}{6} * E(L|X = R) + \frac{1}{3} * E(L|X = B) + \frac{1}{2} * E(L|X = W) \end{aligned}$$

where X is the color of the next roll after the sequence $RBWRB$ has been observed. We then have the following three conditional expectations as follows:

$$E(L|X = x) = \begin{cases} 1 + E(M|R) \text{ if } x = R \\ 1 + E(M|RBW) \text{ if } x = W \\ 1 \text{ if } x = B \end{cases}$$

Substituting these into our expression for $E(L)$, we have:

$$\begin{aligned} E(L) &= \frac{1}{6} * (1 + E(M|R)) + \frac{1}{3} + \frac{1}{2} * (1 + E(M|RBW)) \\ &= 1 + \frac{1}{6}E(M|R) + \frac{1}{2}E(M|RBW) \end{aligned}$$

We can express $E(M|RBW)$ by splitting up the time to obtain the sequence $RBWRBB$ as follows:

$$\begin{aligned} E(M) &= E(\text{time until } RBW) + E(M|RBW) \\ \implies E(M|RBW) &= E(M) - E(t(RBW)) \end{aligned}$$

Similarly for $E(M|R)$:

$$\begin{aligned} E(M) &= E(t(R)) + E(M|R) \\ \implies E(M|R) &= E(M) - E(t(R)) \end{aligned}$$

Substituting these equations into our expression for $E(L)$, we get:

$$\begin{aligned} E(L) &= 1 + \frac{1}{6}E(M|R) + \frac{1}{2}E(M|RBW) \\ &= 1 + \frac{1}{6}(E(M) - E(t(R))) + \frac{1}{2}(E(M) - E(t(RBW))) \\ &= 1 + \frac{2}{3}E(M) - \frac{1}{6}E(t(R)) - \frac{1}{2}E(t(RBW)) \end{aligned}$$

Note that $E(t(R)) = 6$ and $E(t(RBW)) = 6 * 3 * 2$ because both are examples of the case where no overlaps occur. So:

$$\begin{aligned} E(L) &= 1 + \frac{2}{3}E(M) - \frac{1}{6} * 6 - \frac{1}{2} * 6 * 3 * 2 \\ &= \frac{2}{3}E(M) - 18 \\ E(M) - E(N) &= \frac{2}{3}E(M) - 18 \\ 6^2 * 3^3 * 2 - E(N) &= \frac{2}{3}(6^2 * 3^3 * 2) - 18 \end{aligned}$$

Solving this equation, we obtain $E(N) = 234$.

6. Now reconsider tossing the coin which lands heads with probability p and tails with probability $q = 1 - p$. It is continuously flipped until two of the most recent flips are heads. Find the expected number of tosses needed to make this happen.
Solution: Denote N to be the number of flips needed until two of the most recent flips land heads. Also, let X be the time of the first heads.

We can then compute $E(N)$ as follows,

$$E(N) = E(E(N|X))$$

To obtain $E(N|X)$, we shall further condition on the outcomes of the next two flips:

$$\begin{aligned} E(N|X) &= E(N|X,h,h)P(h,h) + E(N|X,h,t)P(h,t) + E(N|X,t,h)P(t,h) + E(N|X,t,t)P(t,t) \\ &= E(N|X,h,h)p^2 + E(N|X,h,t)pq + E(N|X,t,h)qp + E(N|X,t,t)q^2 \end{aligned}$$

Note that:

$$
\begin{aligned}
E(N|X,h,h) &= X+1 \\
E(N|X,h,t) &= X+1 \\
E(N|X,t,h) &= X+2 \\
E(N|X,t,t) &= X+2+E(N)
\end{aligned}
$$

Substituting these into our original expression:

$$
\begin{aligned}
E(N|X) &= (X+1)p^2 + (X+1)pq + (X+2)qp + (X+2+E(N))q^2 \\
&= X + p^2 + 3pq + 2q^2 + E(N)q^2
\end{aligned}
$$

and we have

$$
\begin{aligned}
&E(N) \\
&= E(E(N|X)) \\
&= E(X + p^2 + 3pq + 2q^2 + E(N)q^2) \\
&= E(X) + p^2 + 3pq + 2q^2 + E(N)q^2
\end{aligned}
$$

Note that X is a geometric random variable with parameter p. As such,

$$
\begin{aligned}
&E(X) + p^2 + 3pq + 2q^2 + E(N)q^2 \\
&= \frac{1}{p} + p^2 + 3pq + 2q^2 + E(N)q^2 \\
(1-q^2)E(N) &= \frac{1}{p} + p^2 + 3pq + 2q^2 \\
E(N) &= \frac{\frac{1}{p} + p^2 + 3pq + 2q^2}{1-q^2} \\
&= \frac{1 + p^3 + 3p^2q + 2q^2p}{p(1-q^2)} \\
&= \frac{1 + 2p - p^2}{p(1-q^2)}
\end{aligned}
$$

7. **Polya's Urn Model** Suppose you have an urn that contains r red balls and b blue balls. At each stage, a ball is randomly selected from the urn, then it and m other balls of the same color are put into the urn. Let X_k be the number of red balls among the first k draws.

   (a) Find $E(X_1)$, $E(X_2)$, and $E(X_3)$.
   Solution: In the case where $k=1$, we can condition on the color of the draw as follows:

$$
\begin{aligned}
E(X_1) &= E(X_1|R)P(R) + E(X_1|B)P(B) \\
&= E(X_1|R) * \frac{r}{r+b} + E(X_1|B) * \frac{b}{r+b} \\
&= 1 * \frac{r}{r+b} + 0 * \frac{b}{r+b} \\
&= \frac{r}{r+b}
\end{aligned}
$$

We can make a similar argument for when $k = 2$.

$$\begin{aligned} E(X_2) &= E(X_2|R)P(R) + E(X_2|B)P(B) \\ &= E(X_2|R) * \frac{r}{r+b} + E(X_2|B) * \frac{b}{r+b} \end{aligned}$$

And we have separate expressions for the conditional expectations above.

$$\begin{aligned} E(X_2|R) &= E(X_2|RR) * \frac{r+m}{r+m+b} + E(X_2|RB) * \frac{b}{r+m+b} \\ &= 2 * \frac{r+m}{r+m+b} + 1 * \frac{b}{r+m+b} \\ E(X_2|B) &= E(X_2|BR) * \frac{r}{r+m+b} + E(X_2|BB) * \frac{b+m}{r+m+b} \\ &= 1 * \frac{r}{r+m+b} + 0 * \frac{b+m}{r+m+b} \end{aligned}$$

Plugging this back into our original equation:

$$\begin{aligned} E(X_2) &= E(X_2|R) * \frac{r}{r+b} + E(X_2|B) * \frac{b}{r+b} \\ &= (2 * \frac{r+m}{r+m+b} + 1 * \frac{b}{r+m+b}) * \frac{r}{r+b} + (1 * \frac{r}{r+m+b}) * \frac{b}{r+b} \\ &= \frac{(2(r+m)+b)r + rb}{(r+b)(r+m+b)} \\ &= \frac{2r^2 + 2mr + 2rb}{(r+b)(r+m+b)} \\ &= \frac{2r}{r+b} \end{aligned}$$

For $k = 3$, we will instead use the following approach. First define the indicator random variable as follows:

$$Y_i = \begin{cases} 1 \text{ if } i \text{ is red} \\ 0 \text{ else} \end{cases}$$

Note that we have $E(X_3) = \sum_{i=1}^{3} E(Y_i)$. Then

$$\begin{aligned}
E(Y_3) &= E(E(Y_3|X_2)) \\
&= E(\frac{r + mX_2}{r + 2m + b}) \\
&= \frac{r + mE(X_2)}{r + 2m + b} \\
&= \frac{r + m\frac{2r}{r+b}}{r + 2m + b} \\
&= \frac{r(r + b) + 2mr}{(r + 2m + b)(r + b)} \\
&= \frac{r(r + 2m + b)}{(r + 2m + b)(r + b)} \\
&= \frac{r}{r + b}
\end{aligned}$$

Then $E(X_3) = \frac{3r}{r+b}$

(b) Generalize to find an expression for $E(X_k)$ for any $k \geq 1$.

Solution: In general we have that $E(X_k) = \sum_{i=1}^{k} E(Y_i)$.

We will use induction to derive formulas for $E(Y_j)$ for $j \geq 1$. Suppose we have that $E(X_{i-1}) = \frac{(i-1)r}{r+b}$. Show the same for X_i.

$$\begin{aligned}
E(Y_i) &= E(E(Y_i|X_{i-1})) \\
&= E(\frac{r + mX_{i-1}}{r + (i-1)m + b}) \\
&= \frac{r + mE(X_{i-1})}{r + (i-1)m + b} \\
&= \frac{r + m\frac{(i-1)r}{r+b}}{r + (i-1)m + b} \\
&= \frac{r(r + b) + (i-1)mr}{(r + (i-1)m + b)(r + b)} \\
&= \frac{r(r + (i-1)m + b)}{(r + (i-1)m + b)(r + b)} \\
&= \frac{r}{r + b}
\end{aligned}$$

Hence we have $E(X_i) = \frac{(i-1)r}{r+b} + \frac{r}{r+b} = \frac{ir}{r+b}$.

8. You have two opponents, A and B, with whom you alternate play. Whenever you start with A, you win with probability p_A and when you start with B, you win with probability p_B, where $p_B > p_A$. If your objective is to play the least number of games possible so as to win two games in a row, should you start with A or B?

Solution: Let N_i denote the number of games you play if you start with player i. And let

W_i denote the event that player i wins a game, and $L_i = \overline{W}_i$ be the event that player i loses a game. Then we have the following:

$$\begin{aligned} E(N_A) &= E(N_A|W_A)P(W_A) + E(N_A|L_A)P(L_A) \\ &= (1 + E(N_B|W_A))p_A + (1 + E(N_B))(1 - p_A) \end{aligned}$$

where $E(N_B|W_A)$ is the expected number of additional games starting with player B needed to achieve two wins in a row given that we already have a win:

$$E(N_B|W_A) = 1 * p_B + (1 + E(N_A))(1 - p_B)$$

We can set up a similar equation for the expected number of games played when you start with player B as well:

$$\begin{aligned} E(N_B) &= E(N_B|W_B)P(W_B) + E(N_B|L_B)P(L_B) \\ &= (1 + E(N_A|W_B))p_B + (1 + E(N_A))(1 - p_B) \end{aligned}$$

where $E(N_A|W_B)$ is the expected number of additional games starting with player A needed to achieve two wins in a row given that we already have a win:

$$E(N_A|W_B) = 1 * p_A + (1 + E(N_B))(1 - p_A)$$

9. Two players take turns shooting at a target. Each shot by player i hits the target with probability p_i. All shooting ends when two consecutive shots hit the target. Let m_i denote the mean number of shots taken when player i shoots first.
   (a) Find m_1 and m_2.
   (b) Let h_i denote the mean number of times that the target is hit when player i shoots first. Find h_1 and h_2.

10. Each element in a sequence of binary data is either 1 with probability p or 0 with probability $q = 1 - p$. A maximal subsequence of consecutive values having the same outcomes is called a "run". For instance, if we are given a sequence $1, 1, 1, 0, 0, 1, 1, 1, 1$, then the first run has length 3, the second has length 2, and the third has length 4.
    (a) Find the expected length of the first run.
    Solution: For both parts we will denote N_i to be the length of the ith run. Let X_1 be the first binary number, i.e, X_1 is 1 with probability p and 0 with probability $1 - p$.

    We will condition on X_1 to obtain the expected value of N_1:

$$\begin{aligned} E(N_1) &= E(N_1|X_1 = 1)P(X_1 = 1) + E(N_1|X_1 = 0)P(X_1 = 0) \\ &= E(N_1|X_1 = 1)p + E(N_1|X_1 = 0)q \end{aligned}$$

We will now compute the respective conditional probabilities. List all sequences that start with 1: $(1, 0, \cdots), (1, 1, 0, \cdots), (1, 1, 1, 0, \cdots), \cdots$. They have respective probabilities $pq, p^2q, p^3q, \cdots$. Similarly for sequences that start with 0. Hence:

$$E(N_1|X_1 = 1) = \sum_{i=1}^{\infty} i * p^i q = \frac{qp}{q^2} = \frac{p}{q}$$
$$E(N_1|X_1 = 0) = \sum_{i=1}^{\infty} i * pq^i = \frac{pq}{p^2} = \frac{q}{p}$$

Plugging this into our original expression, we get:

$$\begin{aligned} E(N_1) &= \frac{p^2}{q} + \frac{q^2}{p} \\ &= \frac{p^3 + q^3}{pq} \end{aligned}$$

(b) Find the expected length of the second run.
Solution: We will condition on the first run. Let $\{R_1 = k\}$ denote the event that the first run is all k's, $k = 0, 1$.

If the first run is a sequence of all 1's, then the second run must be a sequence of all 0's; $(1, \cdots, 1, 0, \cdots), (1, \cdots, 1, 0, 0, \cdots), (1, \cdots, 1, 0, 0, 0 \cdots), \cdots$. Similarly for the first run being a sequence of all 0's.

So we have

$$\begin{aligned} E(N_2|R_1 = 1) &= p \sum_{i=1}^{\infty} i * pq^i \\ &= p^2 \sum_{i=1}^{\infty} i * q^k \\ &= p^2 \frac{1}{(1-q)^2} \\ &= 1 \end{aligned}$$

and

$$\begin{aligned} E(N_2|R_1 = 0) &= q \sum_{i=1}^{\infty} i * qp^i \\ &= q^2 \sum_{i=1}^{\infty} i * p^k \\ &= q^2 \frac{1}{(1-p)^2} \\ &= 1 \end{aligned}$$

Applications of Renewal Theory to Patterns:

Let $X_1, X_2, X_3, \cdots$ be iid with $P(X_i = j) = p(j)$ for all $j \geq 0$. Also denote T to be the first time the pattern $s_1, \cdots, s_r$ occurs.

A renewal occurs at time n, $(n \geq r)$
$\Longrightarrow (X_{n-r+1}, \cdots, X_n) = (s_1, \cdots, s_r)$
$\Longrightarrow N(n) =$ number of renewals by time n is a delayed renewal process,
i.e, the distribution of the first arrival time is different from that of all other arrival times.

Recall that by the elementary renewal theorem:

$$\frac{E(N(n))}{n} \to \frac{1}{\mu}$$
$$\frac{Var(N(n))}{n} \to \frac{\sigma^2}{\mu^3}$$

where μ and σ^2 are the mean and variance, respectively, of the time between successive renewals. We will use the above equations to compute these parameters in the case of patterns.

Define

$$I(i) = \begin{cases} 1 \text{ if there is a renewal at time } i \\ 0 \text{ otherwise} \end{cases}$$

Then we have that $P(I(i) = 1) = P(X_{i-r+1} = i_1, \cdots, X_i = i_r) = \prod_{i=1}^{r} p(s_i) =: p$ and

$$N(n) = \sum_{i=r}^{n} I(i)$$
$$E(N(n)) = \sum_{i=1}^{r} E(I(i)) = (n - r + 1)p$$

Especially in the expression for the expectation, we can divide both sides by n and let it go to infinity.

$$E(N(n)) = (n - r + 1)p$$
$$\frac{E(N(n))}{n} = \frac{(n - r + 1)p}{n}$$

and as $n \to \infty$, the left side of the equation goes to $\frac{1}{\mu}$ by the elementary renewal theorem.

$$\frac{1}{\mu} \approx \frac{np}{n} = p$$
$$\mu = \frac{1}{p}$$

For the variance, we have that

$$
\begin{aligned}
Var(N(n)) &= Var(\sum_{i=r}^{n} I(i)) \\
&= \sum_{i=r}^{n} Var(I(i)) + 2\sum_{j=r+1}^{n}\sum_{i<j} Cov(I(i), I(j)) \\
&= (n-r+1)p(1-p) + 2\sum_{j=r+1}^{n}\sum_{i<j} Cov(I(i), I(j))
\end{aligned}
$$

To calculate the covariance between different $I(i)$'s, note that $I(i)$ and $I(j)$ are independent for all i and j more than a distance r apart. Hence

$$Cov(I(i), I(j)) = 0 \text{ for all } |i-j| \geq r$$

So we rewrite $Var(N(n))$:

$$Var(N(n)) = (n-r+1)p(1-p) + 2\sum_{j=1}^{r-1} Cov(I(r), I(r+j))$$

Dividing both sides of the above equation by n, we get:

$$
\begin{aligned}
\frac{Var(N(n))}{n} &= \frac{(n-r+1)p(1-p)}{n} + \frac{2}{n}\sum_{j=r}^{n-1}\sum_{j=1}^{r-1} Cov(I(r), I(r+j)) \\
&= \frac{(n-r+1)p(1-p)}{n} + \frac{2}{n}(n-r+1)\sum_{j=1}^{r-1} Cov(I(r), I(r+j))
\end{aligned}
$$

and as $n \to \infty$, we have:

$$
\begin{aligned}
\frac{\sigma^2}{\mu^3} &= p(1-p) + 2\sum_{j=1}^{r-1} Cov(I(r), I(r+j)) \\
\sigma^2 &= \frac{1-p}{p^2} + \frac{2}{p^3}\sum_{j=1}^{r-1} Cov(I(r), I(r+j))
\end{aligned}
$$

Now, the covariance is dependent on the amount of overlap (the length of the largest pattern at the end of the sequence that is identical to the pattern at the beginning) there is in the sequence. For example, the sequence 00111 has no overlap, whereas the sequence 001100 has an overlap of size 2, from the 00's at the beginning and at the end of the sequence.

In the case where we have no overlap, $N(n)$ becomes just an ordinary renewal process and T is distributed with mean μ and variance σ^2. So we have

$$E(T) = \mu = \frac{1}{p}$$

Because we have no overlap, two patterns cannot be within a distance less than r from each other. Hence, $I(r)I(r+j) = 0$ for all $j \leq r-1$. So

$$\begin{aligned} Cov(I(r), I(r+j)) &= 0 - E(I(r))E(I(r+j)) \\ &= -p^2 \text{ for all } k \leq r-1 \end{aligned}$$

Plugging this into the expression for $Var(T)$, we get

$$\begin{aligned} Var(T) &= \sigma^2 \\ &= \frac{1-p}{p^2} + \frac{2}{p^3}\sum_{j=1}^{r-1} -p^2 \\ &= \frac{1-p}{p^2} - \frac{2(r-1)p^2}{p^3} \\ &= \frac{1}{p^2} - \frac{2r-1}{p} \end{aligned}$$

In the case where we have an overlap of maximum size k in the sequence, we can split it up as follows:

$$T = T_k + A$$

where A is the additional time it takes to get from the sequence in the overlap of size k to the desired original sequence. Then we have that:

$$\begin{aligned} E(T) &= E(T_k) + E(A) \\ Var(T) &= Var(T_k) + Var(A) \end{aligned}$$

Because there is no overlap in the sequence of A, $E(A) = \frac{1}{p}$. Similarly is the case with T_k.

To expand upon the expression of the variance, we first note that no two patterns can occur within a distance of $r-k-1$ from each other. Hence, we have $I(r)I(r+j) = 0$ for all $j \leq r-k-1$.

As such, we get:

$$\begin{aligned} Var(A) &= \sigma^2 \\ &= \frac{1-p}{p^2} + \frac{2}{p^3}\left[\left(\sum_{j=1}^{r-1} E(I(r))E(I(r+j)\right) - (r-1)p^2\right] \\ &= \frac{1}{p^2} - \frac{2r-1}{p} + \frac{2}{p^3}\sum_{j=1}^{r-1} E(I(r))E(I(r+j)) \end{aligned}$$

where $E(I(r))E(I(r+j)$ varies depending on the pattern.

We will now consider the following examples:

1. Determine the number of flips of a fair coin until the pattern $HHTHH$ occurs.

Solution: Note that this sequence has an overlap of maximum size 2 (HH).

In the sequence:

$$\mathrlap{\overbrace{HHT\,HH}^{x}}HHT\,\underbrace{HHTHH}_{y}$$

the overlap is of size 2. The overall length of the sequence is 8. Hence the expected value is given by

$$E\left[I(5)I(8)\right] = P(HHTHHTHH) = \frac{1}{2^8}$$

Similarly, the sequence

$$\mathrlap{\overbrace{HHHT\,H}^{x}}HHHT\,\underbrace{HHTHH}_{y}$$

has an overlap of size 1 and overall sequence length 9. Hence the expected value is given by

$$E\left[I(5)I(9)\right] = P(HHTHHHTHH) = \frac{1}{2^9}$$

We have that $E(A) = 32$. Hence:

$$\begin{aligned} E(T) &= E(T_{HH}) + E(A) \\ &= E(T_{HH}) + 32 \end{aligned}$$

For the pattern HH, we have $p = \frac{1}{4}$ and an overlap of 1. So similar calculations as above leads to:

$$E(T_{H,H}) = E(T_H) + 4$$

So we have that $E(T_k) = 2$. Overall, $E(T) = E(T_k) + 4 + 32 = 38$.

2. Now suppose we have a sequence of numbers from the set $\{0, 1, \cdots, 9\}$ such that $P(X_n = i) = p_i$. Further suppose we desire to calculate the number of trials necessary to observe the pattern 0, 1, 2, 0, 1, 3, 0, 1.
Solution: From our desired pattern above, we have that $p = p^3{}_0 p^3{}_1 p_2 p_3$. Note that the maximum overlap is of size 2. So we have that

$$\begin{aligned} &E\left[I(8)I(14)\right] = p^5{}_0 p^3{}_1 p^2{}_2 p^2{}_3 \\ &E\left[I(8)I(15)\right] = 0 \end{aligned}$$

From the equation $E(T) = E(T_k) + E(A)$, we see that:

$$E(T) = E(T_{0,1}) + \frac{1}{p}$$

Since there is no overlap in the sequence $0, 1$, we get:

$$E(T_{0,1}) = \frac{1}{p_0 p_1}$$

Furthermore, we have

$$Var(T) = Var(T_{0,1}) + \frac{1}{p^2} - \frac{15}{p} + \frac{2}{p(p_0p_1)}$$

Similarly, since there is no overlap in the sequence $0, 1$, we get:

$$Var(T_{0,1}) = \frac{1}{p_0p_1} - \frac{3}{p_0p_1}$$

So overall, we get:

$$\begin{aligned} E(T) &= \frac{1}{p_0p_1} + \frac{1}{p} \\ Var(T) &= \frac{1}{p_0p_1} - \frac{3}{p_0p_1} + \frac{1}{p^2} - \frac{15}{p} + \frac{2}{p(p_0p_1)} \end{aligned}$$

where $p = {p^3}_0{p^3}_1p_2p_3$

3. Continually flip a fair coin. Suppose $A(1) = HTTHH$ and $A(2) = HHTHT$. We want to determine which pattern has a higher chance of occurring first.
Solution: Before we begin this problem, let us introduce a bit of general background.

Suppose there are s disjoint patterns $A(1), A(2), \cdots A(s)$ (disjoint meaning that none of the patterns are contained in any of the others). We are interested in the expected time until one of these patterns occurs and the probability that this pattern occurs first.

Let $T(i)$ denote the time until the pattern $A(i)$ occurs and let $T(i, j)$ denote the additional time that it takes pattern $A(j)$ to occur, starting from the time pattern $A(i)$ occurred, for $i \neq j$.

$E(T(i))$ is computed using the formulas developed previously, in the cases where overlap may or may not occur. As for $E(T(i, j))$, we apply the same procedure, depending on the amount of overlap between the latter part of $A(i)$ and the beginning part of $A(j)$.

For example, consider $A(1) = 1, 1, 3, 0, 1, 2$ and $A(2) = 0, 1, 2, 0, 1$. Then because $T = T_k + A$:

$$T(0) = T_{0,1,2} + T(3, 0)$$

and we get

$$\begin{aligned} E(T(3, 0)) &= E(T(0)) - E(T_{0,1,2}) \\ &= \frac{1}{{p^2}_0{p^2}_1p_2} - \frac{1}{p_1p_0} - \frac{1}{p_0p_1p_2} \end{aligned}$$

because none of the patterns described by each of the terms above have overlaps.

Let $M = \min_i T(i)$ and denote $P(i) = P(M = T(i))$, the probability that $A(i)$ is the first pattern to occur.

For each j, we will derive an equation as follows:

$$\begin{aligned}E(T(j)) &= E(M) + E(T(j) - M)\\ &= E(M) + \sum_{i:i\neq j} E(T(i,j))P(i)\end{aligned}$$

by conditioning on which pattern occurs first.

The above set of equations for all j, along with the constraint that all the $P(j)$'s must sum to 1 give rise to a set of $s+1$ equations in terms of the $s+1$ unknowns $E(M), P(i), i = 1, \cdots s$. So, solving them will yield what we desire to obtain.

Going back to our example, we will merely apply these steps.

We have

$$\begin{aligned}E(T(1)) &= 32 + E(T_h) = 32 + 2 = 34\\ E(T(2)) &= 32\\ E(T(1,2)) &= E(T(2)) - E(T_{h,h}) = 32 - 6 = 26\\ E(T(2,1)) &= E(T(1)) - E(T_{h,t}) = 34 - 4 = 30\end{aligned}$$

We obtain the following set of linear equations:

$$\begin{aligned}34 &= E(M) + 30P(2)\\ 32 &= E(M) + 26P(1)\\ 1 &= P(1) + P(2)\end{aligned}$$

Solving yields $E(M) = 19, P(1) = P(2) = \frac{1}{2}$.

4. Suppose that each play of a game is, independently of the outcomes of the previous plays, won by player i with probability p_i where $i = 1, 2, \cdots, k$. Suppose that there are specific numbers $n(1), n(2), \cdots n(k)$ such that the first player i to win $n(i)$ consecutive plays is declared the winner of the match. Find the expected number of plays until there is a winner and also the probability that the winner is $i, i = 1, 2, \cdots, k$.
   Solution: The previous example set up equations in the case where no overlaps occurred. In this case, we get that

$$E(T(i,j)) = E(T(j))$$

for all $i \neq j$. Then the equations reduce to:

$$\begin{aligned}E(T(j)) &= E(M) + (1 - P(j))E(T(j))\\ P(j) &= \frac{E(M)}{E(T(j))}\end{aligned}$$

Summing the first j equations gives us:

$$E(M) = \left(\sum_{j=1}^{s} \frac{1}{E(T(j))} \right)^{-1}$$

$$P(j) = \frac{1}{E(T(j))} / \sum_{j=1}^{s} \frac{1}{E(T(j))}$$

Applying this to our example, we have

$$E(M) = \left(\sum_{j=1}^{s} \left(\frac{p_j{}^{n(j)}(1-p_j)}{(1-p_j{}^{n(j)})} \right) \right)^{-1}$$

$$P(i) = \frac{p_i{}^{n(i)}(1-p_i)}{(1-p_i{}^{n(i)})} \Big/ \sum_{j=1}^{s} \left(\frac{p_j{}^{n(j)}(1-p_j)}{(1-p_j{}^{n(j)})} \right)$$

This is because if we let $A(i)$ denote the pattern of n_i consecutive values of i, the question becomes a problem of solving for $E(M)$ and $P(i)$ (the probability that the pattern $A(i)$ occurs first).

We have that

$$\begin{aligned} E(T(i)) &= \left(\frac{1}{p_i}\right)^{n(i)} + \left(\frac{1}{p_i}\right)^{n(i)-1} + \cdots + \frac{1}{p_i} \\ &= \frac{1 - p_i{}^{n(i)}}{p_i{}^{n(i)}(1 - p_i)} \end{aligned}$$

As a numerical example, suppose $k = 3$, with player 1 having probability 0.2 of winning, player 2 having probability 0.4 of winning and player 3 having probability 0.6 of winning. Suppose that player 1 needs to win 3 games, player 2 needs to win 5 games and player 3 needs to win 7 games:

$$\begin{aligned} p_1 &= 0.2 \\ p_2 &= 0.4 \\ p_3 &= 0.6 \\ n(1) &= 3 \\ n(2) &= 5 \\ n(3) &= 7 \end{aligned}$$

We want to find the expected number of plays until there is a winner and the probability of winning for each of the three players.

Substituting them into the equations above gives us

$$\begin{aligned} E(M) &= \left(\frac{(0.2)^3(0.8)}{1-(0.2)^3} + \frac{(0.4)^5(0.6)}{1-(0.4)^5} + \frac{(0.6)^7(0.4)}{1-(0.6)^7} \right)^{-1} \\ &= (0.0064516 + 0.0062076 + 0.01152)^{-1} \\ &= 41.36 \end{aligned}$$

and

$$
\begin{aligned}
P(1) &= \frac{0.0064516}{0.0064516 + 0.0062076 + 0.01152} = 0.2668 \\
P(2) &= \frac{0.0062076}{0.0064516 + 0.0062076 + 0.01152} = 0.2567 \\
P(3) &= \frac{0.01152}{0.0064516 + 0.0062076 + 0.01152} = 0.4764
\end{aligned}
$$

Mean Time to a Maximal Run of Different Values

Suppose $X_i, i \geq 1$ are iid random variables that are equally likely to take on any of the values $1, 2, \cdots m$. After being observed sequentially, let T be the first time that a run of m consecutive values includes all the values $1, 2, \cdots m$.

$$
T = \min\{n : X_{n-m+1}, \cdots, X_n \text{ are all distinct}\}
$$

We desire to compute $E(T)$. Define a renewal process where the first renewal occurs at time T, from which you restart and, let the next renewal occur the next time a run of m consecutive values are all different.

Now transform the renewal process into a delayed renewal reward process by assuming that a reward of 1 is earned at time n, for $n \geq m$, if the values $X_{n-m+1}, \cdots, X_n$ are all distinct, i.e, a reward is earned each time the previous m values are all distinct.

Let R_i denote the reward earned at time i. Then by property of renewal reward processes, we have:

$$
\lim_n \frac{E\left(\sum\limits_{i=1}^{n} R_i\right)}{n} = \frac{E(R)}{E(T)}
$$

where R is the rewards earned between renewal periods.

With A_i equal to the set of the first i data values of a renewal run, and B_i to the set of the first i values after this renewal run, we get that:

$$
\begin{aligned}
E(R) &= 1 + \sum_{i=1}^{m-1} E(\text{reward earned at time } i \text{ after a renewal}) \\
&= 1 + \sum_{i=1}^{m-1} P(A_i = B_i) \\
&= 1 + \sum_{i=1}^{m-1} \frac{i!}{m^i} \\
&= \sum_{i=0}^{m-1} \frac{i!}{m^i}
\end{aligned}
$$

Furthermore, for $i \geq m$, we have that

$$\begin{aligned} E(R_i) &= P(X_{i-m+1}, \cdots, X_i \text{ are all different}) \\ &= \frac{m!}{m^m} \end{aligned}$$

Then it follows from the property of renewal reward processes that

$$\frac{m!}{m^m} = \frac{E(R)}{E(T)}$$

and we finally obtain that

$$\begin{aligned} E(T) &= \frac{m^m}{m!} E(R) \\ &= \frac{m^m}{m!} \sum_{i=0}^{m-1} \frac{i!}{m^i} \end{aligned}$$

Now consider the following example: continually flip a coin that lands heads with probability p and tails with probability $q = 1-p$. We shall compute the expected time until the pattern $HHHTHHH$ occurs.

Construct a renewal process by letting the first renewal occur when the pattern first appears, and then start over. Say that a reward of 1 is earned whenever the pattern appears.

If R denotes the reward earned between renewal periods, we have

$$\begin{aligned} E(R) &= 1 + \sum_{i=1}^{6} E(\text{reward earned at time } i \text{ after a renewal}) \\ &= 1 + 0 + 0 + 0 + p^3 q + p^3 qp + p^3 qp^2 \end{aligned}$$

where each sum term after 1 arises from the patterns H, HH, HHH, $HHHT$, $HHHTH$, and $HHHTHH$ respectively.

We have that the expected reward earned at time i is $E(R_i) = p^6 q$ (because the pattern we desire to observe has six heads and one tails)

From the renewal reward property:

$$\begin{aligned} \frac{1 + qp^3 + qp^4 + qp^5}{E(T)} &= qp^6 \\ E(T) &= \frac{1}{qp^6} + \frac{1}{p^3} + \frac{1}{p^2} + \frac{1}{p} \end{aligned}$$

Increasing Runs of Continuous Random Variables

Say we have a sequence of iid random variables $X_1, X_2, \cdots$. Let T denote the first time there is a string of r consecutive increasing values:

$$T = \min\{n \geq r : X_{n-r+1} < X_{n-r+2} < \cdots < X_n\}$$

To compute $E(N(n))$, define a stochastic process whose state at time k, called S_k, is equal to the number of consecutive increasing values starting from time k. That is, for $1 \leq j \leq k$,

$$S_k = j \text{ if } X_{k-j} > X_{k-j+1} < \cdots < X_{k-1} < X_k$$

Note that a renewal occurs at time k iff $S_k = ir$ for some $i \geq 1$. For instance, if $r = 3$ and

$$X_5 > X_6 > \cdots > X_{11}$$

then we have

$$S_6 = 1, S_7 = 2, \cdots, S_{11} = 6$$

We have

$$\begin{aligned} P(S_k = j) &= P(X_{k-j} > X_{k-j+1} < \cdots < X_{k-1} < X_k) \\ &= P(X_{k-j+1} < \cdots < X_{k-1} < X_k) - P(X_{k-j} < X_{k-j+1} < \cdots < X_{k-1} < X_k) \\ &= \frac{1}{j!} - \frac{1}{(j+1)!} \\ &= \frac{j}{(j+1)!} \end{aligned}$$

since all possible orderings of random variables are equally likely.

So we see that

$$\begin{aligned} &\lim_{k \to \infty} P(\text{a renewal occurs at time } k) \\ &= \lim_{k \to \infty} \sum_{i=1}^{\infty} P(S_k = ir) \\ &= \sum_{i=1}^{\infty} \frac{ir}{(ir+1)!} \end{aligned}$$

Using the tail sum formula for expectations, we conclude that

$$E(N(n)) = \sum_{k=1}^{n} P(\text{a renewal occurs at time } k)$$

So we finally obtain that

$$E(T) = \left(\sum_{i=1}^{\infty} \frac{ir}{(ir+1)!} \right)^{-1}$$

Chapter 3: Some Advanced Concepts

Conditional Exp, Martingale, Optional Stopping Thm, copied straight from yellow book...needs total revision

Brownian Motion and Branching Process

To give a concrete example of the kinds of processes we will be examining in this section, we will take a look at a population of individuals, where each individual is able to produce offspring via asexual reproduction (that is, only *one* parent is needed to create offspring and the offspring are just clones of their parent).

Suppose that each individual has probability p_j of having produced j offspring by the end of its lifetime, independent of the other individuals of its population. Denote X_0 to be the number of individuals the population starts out with (called 'the zeroeth generation') and denote X_n to be the number of individuals in the nth generation. For instance, X_1 is the collective number of offspring produced by everyone in the zeroeth generation. Let μ and σ^2 be the mean and variance, respectively, of the number of offspring produced by a single individual. Then:

$$\mu = \sum_{j=0}^{\infty} j p_j$$
$$\sigma^2 = \sum_{j=0}^{\infty} (j-\mu)^2 p_j$$

Any model similar to the one described in the example above is called a branching processes.

Suppose $X_0 = 1$. We ask the following questions:

1. Determine a recursive relationship between the expected number of individuals in the nth generation, and the expected number of individuals in the $n+1$th generation.

   Define N_i to be the number of offspring that person i from the nth generation has. Using the Law of Expectations:

$$\begin{aligned} E(E(X_{n+1}|X_n)) &= E(E(\sum_{i=1}^{X_n} N_i|X_n)) \\ &= E(E(N_1 + \cdots + N_{X_n}|X_n)) \\ &= E(E(N_1|X_n) + ... + E(N_{X_n}|X_n)) \\ &= E(\mu * X_n) \\ &= \mu E(X_n) \end{aligned}$$

   where the second-to-last line comes from the fact that $E(N_i) = \mu$, by definition of N_i.

   We can then use our derived recursive relationship to come up with an exact formula for

$E(X_n)$ for any $n \geq 0$:

$$\begin{aligned} E(X_1) &= \mu E(X_0) = \mu \\ E(X_2) &= \mu E(X_1) = \mu^2 \\ &\vdots \\ E(X_n) &= \mu^n \end{aligned}$$

2. Repeat the previous question with the variances instead.

We can use the Law of Total Variance:

$$Var(X_{n+1}) = E(Var(X_{n+1}|X_n)) + Var(E(X_{n+1}|X_n))$$

For the first term, we are given that there are X_n individuals in the previous generation. Each individual will have variance σ^2 in the number of offspring it produces. Hence:

$$\begin{aligned} E(Var(X_{n+1}|X_n)) &= E(\sigma^2 * X_n) \\ &= \sigma^2 E(X_n) \\ &= \sigma^2 \mu^n \end{aligned}$$

We can use the previous question to simplify the second term:

$$\begin{aligned} E(Var(X_{n+1}|X_n)) &= E(Var(\mu X_n)) \\ &= \mu^2 Var(X_n) \end{aligned}$$

Overall,

$$Var(X_{n+1}) = \sigma^2 \mu^n + \mu^2 Var(X_n)$$

We can similarly derive a definite expression from the recursive relationship above. First, we split things into two cases: when $\mu = 1$ and $\mu > 1$.

When $\mu = 1$, we have

$$\begin{aligned} Var(X_{n+1}) &= \sigma^2 + Var(X_n) \\ \implies Var(X_1) &= \sigma^2 + Var(X_0) = \sigma^2 \\ Var(X_2) &= \sigma^2 + Var(X_1) = 2\sigma^2 \\ &\vdots \\ Var(X_n) &= n\sigma^2 \end{aligned}$$

When $\mu > 1$, we have

$$\begin{aligned}
Var(X_1) &= \sigma^2\mu^0 + \mu^2 Var(X_0) = \sigma^2 \\
Var(X_2) &= \sigma^2\mu + \mu^2 Var(X_1) \\
&= \sigma^2\mu + \mu^2\sigma^2 \\
&= \sigma^2\mu(1+\mu) \\
Var(X_3) &= \sigma^2\mu^2 + \mu^2(\sigma^2\mu + \mu^2\sigma^2) \\
&= \sigma^2\mu^2 + \mu^3\sigma^2 + \mu^4\sigma^2 \\
&= \sigma^2\mu^2(1+\mu+\mu^2) \\
&\vdots
\end{aligned}$$

and the general formula

$$Var(X_n) = \sigma^2\mu^{n-1}\left(\frac{\mu^n - 1}{\mu - 1}\right)$$

3. Let π denote the probability that a population that starts with one individual will eventually die out. Argue that π satisfies the following expression

$$\pi = \sum_{j=0}^{\infty} \pi^j p_j$$

We can solve this problem by conditioning on the number of offspring of the initial member of generation n.

$$\begin{aligned}
\pi &= P(\text{population dies out}) \\
&= \sum_j P(\text{population dies out}|X_n = j)p_j \\
&= \sum_j \pi^j p_j
\end{aligned}$$

where the last line comes from the fact that each of the j individuals of the ith generation can be thought of as starting their own independent branching process.

additional extensions...rewriting above as Markov chain with different probability models of offspring. Verifying Martingale property with this model, etc.

Now consider the symmetric random walk, which in each time unit, is equally likely to take a unit step up or down. Suppose we take smaller and smaller time steps and eventually take a limit. What we obtain is known as a Brownian Motion.

To reiterate the above, let Δt denote the length of the time unit we take and Δx denote the length of the step size we take per time unit.

Further define the indicator random variable:

$$X_i = \begin{cases} +1 \text{ if going up at } ith \text{ step} \\ -1 \text{ if going down at } ith \text{ step} \end{cases}$$

and $P(X_i = 1) = P(X_i = -1) = \frac{1}{2}$.

Now let $X(t)$ denote the overall position we are at in time t:

$$X(t) = \Delta x(X_1 + \cdots + X_{\frac{t}{\Delta t}})$$

We have the expected value:

$$\begin{aligned}
E(X(t)) &= E(\Delta x(X_1 + \cdots + X_{\frac{t}{\Delta t}})) \\
&= \Delta x E(X_1 + \cdots + X_{\frac{t}{\Delta t}}) \\
&= \frac{t}{\Delta t}\Delta x E(X_i) \\
&= 0
\end{aligned}$$

and the variance

$$\begin{aligned}
Var(X(t)) &= Var(\Delta x(X_1 + \cdots + X_{\frac{t}{\Delta t}})) \\
&= (\Delta x)^2 Var(X_1 + \cdots + X_{\frac{t}{\Delta t}}) \\
&= \frac{t}{\Delta t}(\Delta x)^2 Var(X_i) \\
&= \frac{t}{\Delta t}(\Delta x)^2 E({X_i}^2) \\
&= \frac{t}{\Delta t}(\Delta x)^2
\end{aligned}$$

As we are taking the limit, we must essentially let Δx and Δt go to zero. Let us take $\Delta x = \sigma\sqrt{\Delta t}$ for some $\sigma > 0$ and let Δt go to zero. Then we obtain

$$\begin{gathered}
E(X(t)) = 0 \\
Var(X(t)) \to \sigma^2 t
\end{gathered}$$

Furthermore, by the central limit theorem,

$$P\left(\frac{X(t) - 0}{\sigma^2 t} > x\right) \approx \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-a^2/2} da$$

so

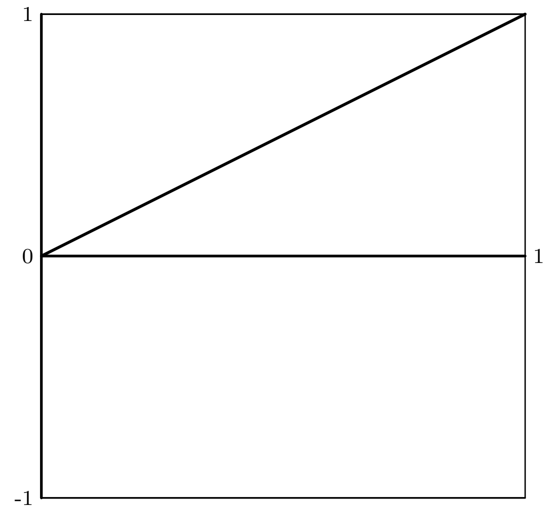
$$\frac{X(t)}{\sigma^2 t} \to \mathcal{N}(0,1)$$

and $X(t)$ is normally distributed with mean 0 and variance $\sigma^2 t$.

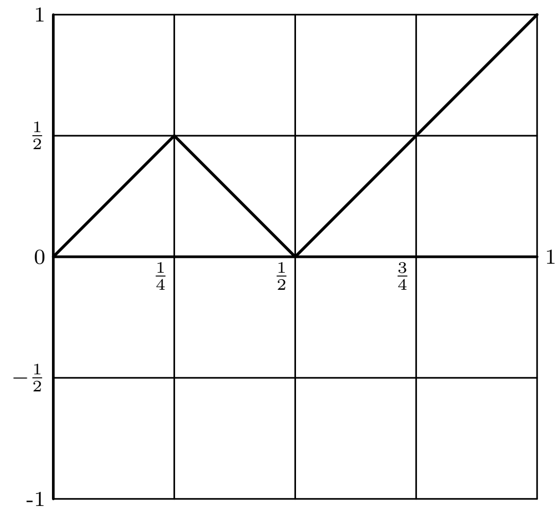
As a more intuitive restatement of what we did above, consider the following sequence of coin toss results:

$$HTHHTHTTHTTTTHHTH$$

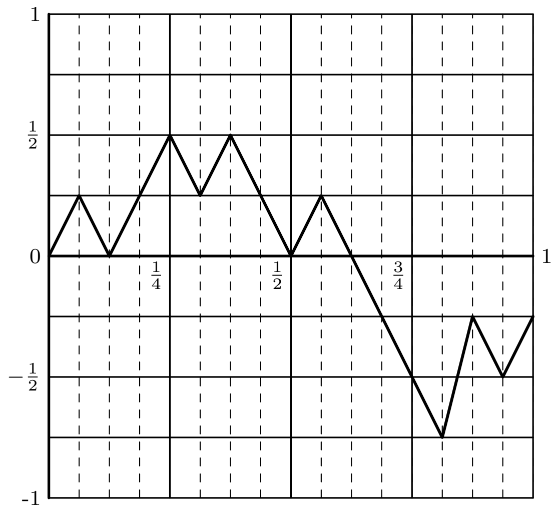
We first start out with



Because when the length interval splits by a certain amount, the time interval splits squared as much, we split the time interval by 4 once we split the length interval by 2. Now we plot the random walk sequence of the first four coin toss results:



Similarly, splitting the length interval by 4 splits the time interval by 16, and we can plot the whole sequence of coin tosses:



Our mathematical analysis above claims that repeating this splitting process and following some sample sequence of an infinite number of coin tosses, e.g, $HTHHTHTTHTTTTHHTH\cdots$, converges to a Brownian motion. Of course, there are many sample paths of coin tosses we can do this splitting process on. In fact, there are 2^∞ such sample paths, an uncountable amount (despite the fact that the rational numbers are a countable set, by Cantor; I find this quite intriguing because it seems that the number of elements in $\mathbb{Q}$ would be much larger than 2^∞. We can prove these facts using real analysis, but the proofs have been omitted because they are not the main point of this text.)
It is also interesting to consider the speed of the Brownian motion process:

Claim: The *speed* of the particle undergoing Brownian motion, with equal probability of going up or down, is infinity.

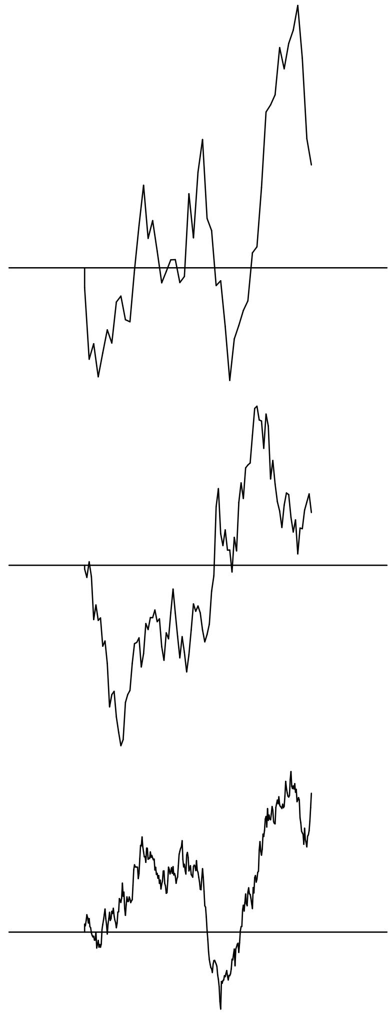
Proof: At each time length of Δt, we either move up or down a vertical distance of Δx. This implies that we actually travel a distance of $\sqrt{\Delta x^2 + \Delta t^2}$ at each time length Δt.

We have seen before that $\Delta x = \sqrt{\Delta t}$, assuming a variance of 1. Measuring up to a total time of t, there are $\frac{t}{\Delta t}$ time steps.

Putting all this info together, the speed is given by:

$$\begin{aligned}
&\frac{\sqrt{\Delta x^2 + \Delta t^2} * \frac{t}{\Delta t}}{t} \\
&= \frac{\sqrt{\Delta t + \Delta t^2} * \frac{t}{\Delta t}}{t} \\
&= \frac{\sqrt{\Delta t}\sqrt{1 + \Delta t}}{\Delta t} \\
&= \sqrt{\frac{1 + \Delta t}{\Delta t}} \\
&= \sqrt{1 + \frac{1}{\Delta t}}
\end{aligned}$$

We are interested in the case where Δt tends to zero, as in the figures below for Brownian Motion ($\Delta t \to 0$ means that the movement for each step becomes more and more rapid):



But note that as $\Delta t \to 0$, the expression above tends to infinity. This proves our claim.

We shall now consider some properties of the Brownian motion process:

1. $X(t)$ is normal with mean 0 and variance $\sigma^2 t$ because of the central limit theorem.
2. $\{X(t), t \geq 0\}$ has independent increments

$$X(t_n) - X(t_{n-1}), X(t_{n-1}) - X(t_{n-2}), \cdots X(t_2) - X(t_1), X(t_1)$$

   for $t_1 < t_2 < \cdots t_n$. This is because the changes in value of the random walk in nonoverlapping time intervals are independent.
3. $\{X(t), t \geq 0\}$ has stationary increments, i.e, the distribution of $X(t+s) - X(t)$ does not depend on t. This is because the distribution of the change in position of the random walk over any time only depends on the length of the interval, and not when it occurred.

When $\sigma = 1$, the process is a *standard Brownian Motion process.* We will denote this by $\{B(t), t \geq 0\}$.

We say that $\{X(t), t \geq 0\}$ is a Brownian Motion process with a drift coefficient of μ and variance parameter σ^2 if

1. $X(0) = 0$
2. $\{X(t), t \geq 0\}$ has stationary and independent increments
3. $X(t)$ is normally distributed with mean μt and variance $\sigma^2 t$

We can also use the formulation

$$X(t) = \sigma B(t) + \mu t$$

where, again, $B(t)$ is the standard Brownian Motion process.

We now consider the following problem: a variation of the analysis we did in the proof above and an extension to the Gambler's Ruin problem we've analyzed in the beginning of this text.

Suppose we consider the same scenario, except that we either go vertically up with probability p and down with probability $1 = p$ and $p = \frac{1}{2}(1 + \mu\sqrt{\Delta t})$.

1. First, argue that as $\Delta t \to 0$, the resulting process is a Brownian Motion process with drift rate μ.
   Solution: Define

$$X_i = \begin{cases} +1 \text{ if going up at } ith \text{ step} \\ -1 \text{ if going down at } ith \text{ step} \end{cases}$$

   Further define $X(t)$ to be the overall position at time t. Then the following expression is true:

$$X(t) = \sqrt{\Delta t} \sum_{i=1}^{\frac{t}{\Delta t}} X_i$$

   Then we get

$$\begin{aligned} E(X(t)) &= E(\sqrt{\Delta t} \sum_{i=1}^{\frac{t}{\Delta t}} X_i) \\ &= \sqrt{\Delta t} \sum_{i=1}^{\frac{t}{\Delta t}} E(X_i) \\ &= \sqrt{\Delta t} * \frac{t}{\Delta t} * E(X_i) \\ &= \sqrt{\Delta t} * (\frac{t}{\Delta t}) * \mu\sqrt{\Delta t} \end{aligned}$$

   because

$$\begin{aligned} E(X_i) &= 1 * p + (-1) * (1 - p) \\ &= 2p - 1 \\ &= \mu\sqrt{\Delta t} \end{aligned}$$

Similarly for the variance:

$$\begin{aligned} Var(X(t)) &= Var(\sqrt{\Delta t}\sum_{i=1}^{\frac{t}{\Delta t}} X_i) \\ &= \Delta t \sum_{i=1}^{\frac{t}{\Delta t}} Var(X_i) \\ &= \Delta t * \frac{t}{\Delta t} * Var(X_i) \\ &= \Delta t * \frac{t}{\Delta t} * (1-\mu^2\Delta t) \end{aligned}$$

because

$$\begin{aligned} Var(X_i) &= E({X_i}^2) - (E(X_i))^2 \\ &= 1-\mu^2\Delta t \end{aligned}$$

since ${X_i}^2 = 1$ with probability 1.

Now note that as $\Delta t \to 0$, $E(X(t)) \to \mu t$ and $Var(X(t)) \to t$. This is exactly a Brownian motion with a drift rate of μ.

2. Using the previous question and the results of Gambler's Ruin, determine the probability that a Brownian Motion process with drift rate μ goes up a before going down b, where $a, b > 0$.
   Solution: Recall we derived that the probability of going up a before going down b to be

$$\frac{1-\left(\frac{q}{p}\right)^b}{1-\left(\frac{q}{p}\right)^{a+b}}$$

   where p is the probability of going up in one round of a game and $q = 1-p$ is that of going down.

   We are given that $p = \frac{1}{2}(1+\mu\sqrt{\Delta t})$. So

$$\begin{aligned} q = 1-p &= 1-\tfrac{1}{2}(1+\mu\sqrt{\Delta t}) \\ &= \tfrac{1}{2}(1-\mu\sqrt{\Delta t}) \\ \implies \tfrac{q}{p} &= \frac{1-\mu\sqrt{\Delta t}}{1+\mu\sqrt{\Delta t}} \end{aligned}$$

   Hence, in this case, (we must divide each length by $\sqrt{\Delta t}$ because that is the size of the step) the probability of going up $\frac{a}{\sqrt{\Delta t}}$ before going down $\frac{b}{\sqrt{\Delta t}}$ is as follows:

$$\frac{1-\left(\frac{1-\mu\sqrt{\Delta t}}{1+\mu\sqrt{\Delta t}}\right)^{\frac{b}{\sqrt{\Delta t}}}}{1-\left(\frac{1-\mu\sqrt{\Delta t}}{1+\mu\sqrt{\Delta t}}\right)^{\left(\frac{a+b}{\sqrt{\Delta t}}\right)}}$$

Taking the limit of the above as Δt goes to zero. First we will consider the following limit:

$$\begin{aligned}
&\lim_{\Delta t\to 0}(\frac{1-\mu\sqrt{\Delta t}}{1+\mu\sqrt{\Delta t}})^{\frac{1}{\sqrt{\Delta t}}}\\
&=\lim_{h\to 0}(\frac{1-\mu h}{1+\mu h})^{\frac{1}{h}}\\
&=\lim_{g\to\infty}(\frac{1-\frac{\mu}{g}}{1+\frac{\mu}{g}})^{g}\\
&=\frac{e^{-\mu}}{e^{\mu}}\\
&=e^{-2\mu}
\end{aligned}$$

since

$$\lim_{n\to\infty}(1+\tfrac{x}{n})^n=e^x$$

Now, plugging this into our original expression as $\Delta t\to 0$, we get

$$\frac{1-e^{-2\mu b}}{1+e^{-2\mu(a+b)}}$$

We can further consider the distribution of the standard Brownian motion process. $B(t)$ is normal with mean 0 and variance t:

$$f_t(x)=\tfrac{1}{\sqrt{2\pi t}}e^{\frac{-x^2}{2t}}$$

We can also compute an expression for the joint density function $X(t_1), X(t_2), \cdots, X(t_n)$ for $t_1, t_2, \cdots, t_n$. Describe the following set of equations:

$$\begin{aligned}
X(t_1)&=x_1\\
X(t_2)&=x_2\\
&\vdots\\
X(t_n)&=x_n
\end{aligned}$$

Note that these are equivalent to:

$$\begin{aligned}
X(t_1)&=x_1\\
X(t_2)-X(t_1)&=x_2-x_1\\
X(t_3)-X(t_2)&=x_3-x_2\\
&\vdots\\
X(t_n)-X(t_{n-1})&=x_n-x_{n-1}
\end{aligned}$$

As a property of the Brownian Motion process, the in-between time intervals are independent. Hence we can express the joint density as follows:

$$\begin{aligned}
&f(x_1,x_2,\cdots,x_n)\\
&=f_{t_1}(x_1)f_{t_2-t_1}(x_2-x_1)f_{t_3-t_2}(x_3-x_2)\cdots f_{t_n-t_{n-1}}(x_n-x_{n-1})
\end{aligned}$$

$$= \frac{\exp(-\frac{1}{2}(\frac{x_1{}^2}{t_1}+\frac{x_2-x_1{}^2}{t_2-t_1}+\cdots\frac{x_n-x_{n-1}{}^2}{t_n-t_{n-1}}))}{(2\pi)^{\frac{n}{2}}(t_1(t_2-t_1)\cdots(t_n-t_{n-1}))^{\frac{1}{2}}}$$

Given this expression, we can compute the conditional density of $X(s)$ given that $X(t) = a$ where $s < t$:

$$\begin{aligned}
f_{s|t}(x|a) &= \frac{f_s(x)f_{t-s}(a-x)}{f_t(a)} \\
&= K_1 \exp\left[-\frac{x^2}{2s} - \frac{(a-x)^2}{2(t-s)}\right] \\
&= K_1 \exp\left[-\frac{x^2}{2s} - \frac{a^2-2ax+x^2}{2(t-s)}\right] \\
&= K_2 \exp\left[-x^2(\frac{1}{2s} + \frac{1}{2(t-s)}) - \frac{-2ax}{2(t-s)}\right] \\
&= K_2 \exp\left[-x^2(\frac{1}{2s} + \frac{1}{2(t-s)}) + \frac{ax}{(t-s)}\right] \\
&= K_2 \exp\left[-x^2(\frac{2t}{2s(2(t-s))}) + \frac{ax}{(t-s)}\right] \\
&= K_2 \exp\left[-x^2(\frac{t}{2s(t-s))}) + \frac{ax}{(t-s)}\right] \\
&= K_2 \exp\left[-\frac{t}{2s(t-s)}(x^2 - 2\frac{as}{t}x)\right]
\end{aligned}$$

$$= K_3 \exp\left[\frac{(x-\frac{as}{t})^2}{\frac{2s(t-s)}{t}}\right]$$

where K_1, K_2, K_3 are constants. This means that the conditional distribution of $X(s)$ given $X(t) = a$ for $s < t$ is normal with

$$E(X(s)|X(t) = a) = \tfrac{s}{t}a$$
$$Var(X(s)|X(t) = a) = \tfrac{s}{t}(t-s)$$

Now let T_a denote the first time the Brownian Motion process hits value a. Note that when $a < 0$, the distribution of T_a is the same as that of T_{-a} by symmetry. So if the probability of going up in one time step was a generic probability $p \neq \frac{1}{2}$, then the distributions would be different. Hence, we need only consider the case where $a > 0$.

For $a > 0$, we shall compute $P(T_a \leq t)$ by considering $P(X(t) \geq a)$ and conditioning on the event that $T_a \leq t$:

$$\begin{aligned}
P(X(t) \geq a) &= P(X(t) \geq a|T_a \leq t)P(T_a \leq t) + P(X(t) \geq a|T_a > t)P(T_a > t) \\
&= \frac{1}{2} * P(T_a \leq t) + 0 * P(T_a > t)
\end{aligned}$$

since if $T_a \leq t$ then the process hits a at some time less than or equal to t and from there, it is equally likely to be above or below a at time t. The second term probability is 0 because the value cannot be larger than a without having hit a before.

Now we can solve for our desired quantity:

$$\begin{aligned}
P(T_a \leq t) &= 2P(X(t) \geq a) \\
&= \frac{2}{\sqrt{2\pi t}} \int_a^\infty \exp\{\frac{-x^2}{2t}\}dx
\end{aligned}$$

Using the change of variables $y = \frac{x}{\sqrt{t}}$,

$$\begin{aligned}
&\frac{2}{\sqrt{2\pi t}} \int_a^\infty \exp\{\frac{-x^2}{2t}\}dx \\
&= \frac{2}{\sqrt{2\pi}} \int_{\frac{a}{\sqrt{t}}}^\infty \exp\{\frac{-y^2}{2}\}dy
\end{aligned}$$

And similarly, for $a < 0$, we get

$$P(T_a \leq t) = \frac{2}{\sqrt{\pi}} \int_{\frac{|a|}{\sqrt{t}}}^\infty \exp\{\frac{-y^2}{2}\}dy$$

We can also consider the distribution of the maximum value that the process reaches in the time interval $[0, t]$. For $a > 0$:

$$\begin{aligned}
P(\max_{0 \leq s \leq t} X(s) \geq a) &= P(T_a \leq t) \\
&= 2P(X(t) \geq a) \\
&= \frac{2}{\sqrt{2\pi}} \int_{\frac{a}{\sqrt{t}}}^\infty \exp\{\frac{-y^2}{2}\}dy
\end{aligned}$$

And we can also consider the probability of the Brownian motion hitting value a before $-b$, where $a, b > 0$. This is the same as when we previously considered the Gambler's Ruin problem for the symmetric case:

$$P(\text{up } a \text{ before down } b) = \tfrac{b}{a+b}$$

A special case of the Brownian motion process, called the *Geometric Brownian Motion Process* is defined as follows: let $\{X(t), t \geq 0\}$ be a Brownian Motion process with a drift coefficient of μ and variance parameter σ^2. Then the process $\{Y(t), t \geq 0\}$ such that

$$Y(t) = e^{X(t)}$$

Let us compute the expected value of the process at time t given the history or the process up to time s. Then

$$\begin{aligned}
E(Y(t)|Y(u), 0 \leq u \leq s) &= E(e^{X(t)}|X(u), 0 \leq u \leq s) \\
&= E(e^{X(s)+(X(t)-X(s))}|X(u), 0 \leq u \leq s) \\
&= e^{X(s)}E(e^{X(t)-X(s)}|X(u), 0 \leq u \leq s) \\
&= e^{X(s)}E(e^{X(t)-X(s)})
\end{aligned}$$

where the last equality comes from the fact that the time increments are independent.

Recall that the moment generating function of a normal random variable N is given by

$$E(aN) = e^{aE(N)+a^2Var(N)/2}$$

So we have

$$\begin{aligned} E(e^{X(t)-X(s)}) &= e^{\mu(t-s)+(t-s)\sigma^2/2} \\ E(Y(t)|Y(u), 0 \le u \le s) &= e^{X(s)}e^{\mu(t-s)+(t-s)\sigma^2/2} \\ &= Y(s)e^{(t-s)(\mu+\sigma^2/2)} \end{aligned}$$

More Brownian Motion Problems:

1. What is the distribution of $B(s) + B(t)$ for $s \le t$?
   Solution: We can rewrite the sum as follows:

   $$\begin{aligned} B(s) + B(t) &= B(s) + (B(t) - B(s)) + B(s) \\ &= (B(t) - B(s)) + 2B(s) \end{aligned}$$

   Note that $B(s)$ and $B(t) - B(s)$ are independent increments of the Brownian motion process, both normally distributed:

   $$\begin{gathered} 2B(s) \sim N(0, 4s) \\ B(t) - B(s) \sim N(0, t-s) \end{gathered}$$

   Therefore, $(B(t) - B(s)) + 2B(s) \sim N(0, 4s + (t-s)) = N(0, 3s + t)$.

2. Compute the conditional distribution of $B(s)$ given that $B(t_1) = a$ and $B(t_2) = b$ for $0 < t_1 < s < t_2$.
   Solution: The conditional distribution of $B(s)$ given $B(t_1) = a$ and $B(t_2) = b$ is the same as the conditional distribution of $B(s) - a$ given $B(t_1) = a$ and $B(t_2) = b$. This in turn, is equivalent to the conditional distribution of $B(s - t_1)$ given $B(0) = 0$ and $B(t_2 - t_1) = b - a$, by a shift in the axes t_1 to the right and up a units.

   Recall the density for conditional distributions like this: in this case, $B(s - t_1)$ given the shifted conditions above is normally distributed with mean

   $$\tfrac{s-t_1}{t_2-t_1}(b-a)$$

   and variance

   $$\tfrac{s-t_1}{t_2-t_1}(t_2-s)$$

   Now we must shift back to our original scale for our final distribution: $B(s)$ given the original conditions above is normally distributed with mean

   $$a + \tfrac{s-t_1}{t_2-t_1}(b-a)$$

   and variance

   $$\tfrac{s-t_1}{t_2-t_1}(t_2-s)$$

3. Show that the standard Brownian motion is a Martingale.
Solution:

$$\begin{aligned}
E(B(t)|B(s)) &= E(B(s) + (B(t) - B(s))|B(s)) \\
&= B(s) + E((B(t) - B(s))|B(s)) \\
&= B(s) + E(B(t) - B(s))
\end{aligned}$$

since $B(t)-B(s)$ is an independent time interval from $B(s)$. Now note that $E(B(t)-B(s)) = 0$, as characteristic of the standard Brownian motion process. Hence

$$E(B(t)|B(s)) = B(s)$$

which verifies the Martingale property.

4. Compute $E(B(t_1)B(t_2)B(t_3))$ for $t_1 < t_2 < t_3$.
Solution:

$$\begin{aligned}
E(B(t_1)B(t_2)B(t_3)) &= E(E(B(t_1)B(t_2)B(t_3))|B(t_1)B(t_2)) \\
&= E(B(t_1)B(t_2)B(t_2))
\end{aligned}$$

because the standard Brownian motion process is a Martingale.

$$\begin{aligned}
E(B(t_1)B(t_2)^2) &= E(E(B(t_1)B(t_2)^2)|B(t_1)) \\
&= E(B(t_1)E(B(t_2)^2)|B(t_1))
\end{aligned}$$

Note that given $B(t_1)$, $B(t_2)$ is normal with mean $B(t_1)$ and variance $t_2 - t_1$.

$$\begin{aligned}
t_2 - t_1 &= E(B(t_2)^2|B(t_1)) - E(B(t_2)|B(t_1))^2 \\
&= E(B(t_2)^2|B(t_1)) - B(t_1)^2 \\
E(B(t_2)^2|B(t_1)) &= B(t_1)^2 + (t_2 - t_1)
\end{aligned}$$

Continuing our previous calculations:

$$\begin{aligned}
E(B(t_1)E(B(t_2)^2)|B(t_1)) &= E(B(t_1)(B(t_1)^2 + (t_2 - t_1))) \\
&= E(B(t_1)^3) + (t_2 - t_1)E(B(t_1)) \\
&= 0
\end{aligned}$$

5. Show that

$$\begin{aligned}
P(T_a < \infty) &= 1 \\
E(T_a) &= \infty \text{ for } a \neq 0
\end{aligned}$$

Solution: We can express:

$$\{T_a < \infty\} = \lim_{t\to\infty} P(T_a \leq t)$$

Then

$$\begin{aligned}\lim_{t\to\infty} P(T_a \le t) &= \frac{2}{\sqrt{2\pi}}\int_0^\infty e^{-\frac{y^2}{2}}dy \\ &= 2 * P(N(0,1) \ge 0) \\ &= 2 * \frac{1}{2} \\ &= 1\end{aligned}$$

Hence $P(T_a < \infty) = 1$.

Recall $E(T_a) = \int_0^\infty P(T_a > t)dt$. Using this:

$$\begin{aligned}\int_0^\infty P(T_a > t)dt &= \int_0^\infty (1 - P(T_a \le t))dt \\ &= \int_0^\infty dt - \int_0^\infty P(T_a \le t)dt\end{aligned}$$

Note that although the second term (which denotes the area under a scaled normal curve) is bounded and less than infinity, the first term in infinite.

Hence $E(T_a) = \infty$.

6. What is $P(T_2 < T_{-1} < T_4)$?
Solution: We calculate:

$$\begin{aligned}P(T_2 < T_{-1} < T_4) &= P(\text{hit 2 before -1 before 4}) \\ &= P(\text{hit -1 before 4}|\text{hit 2 before -1}) * P(\text{hit 2 before -1})\end{aligned}$$

Note that when starting from zero and considering a standard Brownian motion process, the probability of going up to $+2$ is half as likely (because twice the distance) as going down one to -1:

$$P(\text{hit 2 before -1}) = \tfrac{1}{3}$$

and the probability $P(\text{hit -1 before 4}|\text{hit 2 before -1})$ is the same as considering the probability of reaching -1 before $+4$, given that you start at position $+2$. In the scenario of the symmetric standard Brownian motion, this occurs with probability $\frac{2}{5}$.

$$P(\text{hit -1 before 2}|\text{hit 1 before -1}) = \tfrac{2}{5}$$

Overall:

$$\begin{aligned}&P(\text{hit -1 before 2}|\text{hit 1 before -1}) * P(\text{hit 1 before -1}) \\ &= \frac{1}{3} * \frac{2}{5} \\ &= \frac{2}{15}\end{aligned}$$

Exercise: Consider a generalization of the above problem and determine the probability of

$$A = \{T_1 < T_{-1} < T_2 < T_{-2} < T_3 < T_{-3} < \cdots T_n < T_{-n}\}$$

for any $n > 0$, given that you start from 0.
Solution: We keep conditioning on the past events: $\{T_1 < T_{-1} < T_2 < T_{-2} < T_3 < T_{-3} < \cdots T_n < T_{-n}\}$

$$\begin{aligned}
P(A) &= P(T_1 < T_{-1})P(T_{-1} < T_2|T_1 < T_{-1})P(T_2 < T_{-2}|T_1 < T_{-1} < T_2) \\
&\quad \cdots P(T_n < T_{-n}|T_1 < T_{-1} < T_2 < T_{-2} < T_3 < T_{-3} < \cdots T_{n-1}) \\
&= \frac{1}{2} * \frac{1}{3} * \frac{1}{4} * \cdots * \frac{1}{2n}
\end{aligned}$$

because each conditioned event $\{T_k < T_{-k}|T_1 < T_{-1} < T_2 < T_{-2} < T_3 < T_{-3} < \cdots T_k\}$ is like the event that you hit value k before $-k$, given that you start at value $k-1$. Because the distance from $k-1$ to $-k$ is $2k-1$ times larger than the distance between $k-1$ and k, the probability of reaching $-k$ before k, starting at $k-1$ is $\frac{1}{2k}$.

7. Let $A = \{\max_{t_1 \le s \le t_2} B(s) > x\}$. Compute an expression for $P(A)$.
Solution: Using the formula for conditional probability:

$$\begin{aligned}
P(A) &= \int_{-\infty}^{\infty} P(A|B(t_1) = y) f_{B(t_1)}(y) dy \\
&= \int_{-\infty}^{\infty} P(A|B(t_1) = y) \frac{1}{\sqrt{2\pi t_1}} e^{-\frac{y^2}{2}} dy
\end{aligned}$$

It is then easier to find expressions for $P(A|B(t_1) = y)$. If $y \ge x$, then this probability is 1.

If $y < x$, then:

$$\begin{aligned}
P(A|B(t_1) = y) &= P\left(\max_{t_1 \le s \le t_2} B(s) > x \middle| B(t_1) = y\right) \\
&= P\left(\max_{0 \le s \le t_2 - t_1} B(s) > x - y\right)
\end{aligned}$$

by shifting the whole process back by time length t_1.

8. Let $\{X(t), t \ge 0\}$ be a Brownian Motion process with drift coefficient μ and variance parameter σ^2. What is the joint density function of $X(s)$ and $X(t), s < t$?
Solution: We have that

$$\begin{aligned}
X(t) &= \sigma B(t) + \mu t \\
X(s) &= \sigma B(s) + \mu s
\end{aligned}$$

Using the fact that Brownian motion processes have <u>independent</u> increments, the joint density is described as follows:

$$f(x_s, x_t) = f_s(x_s) f_{t-s}(x_t - x_s)$$

Recall that the density is Gaussian. Hence

$$f_s(x_s) = \frac{1}{\sqrt{2\pi\sigma^2 s}} e^{-\frac{(x_s - \mu s)^2}{2\sigma^2 s}}$$

and

$$f_{t-s}(x_t - x_s) = \frac{1}{\sqrt{2\pi\sigma^2(t-s)}} e^{-\frac{((x_t - x_s) - \mu(t-s))^2}{2\sigma^2(t-s)}}$$

A multiplication of these two densities gives us our desired result.

9. If $\{Y(t), t \geq 0\}$ is a Martingale, show that

$$E(Y(t)) = E(Y(0))$$

Solution: Recall that a stochastic process $\{Y(t), t \geq 0\}$ is a martingale process if for $s < t$,

$$E(Y(t)|Y(u), 0 \leq u \leq s) = Y(s)$$

Let $s = 0$. Since $Y(t)$ is a martingale,

$$\begin{aligned} &E(Y(t)|Y(u), 0 \leq u \leq s) = Y(s) \\ \implies &E(E(Y(t)|Y(u), 0 \leq u \leq s)) = E(Y(s)) \end{aligned}$$

The left side of the equation equals $E(Y(t))$ by the law of total expectation. Hence $E(Y(t)) = E(Y(0))$.

10. Show that $\{Y(t), t \geq 0\}$ is a Martingale when

$$Y(t) = B(t)^2 - t$$

(Recall we showed in the previous chapter that ${Y_n}^2 - n$ is a Martingale, where Y_n was a random walk process.) Furthermore, what is $E(Y(t))$?
Solution: Let us slightly redefine the above notion of a Martingale: a stochastic process $\{Y(t), t \geq 0\}$ is a martingale process if for $s < t$,

$$E(Y(t)|\mathcal{F}_s) = Y(s)$$

that is, $Y(t)$ is $\mathcal{F}_s$-measurable.

In the context of this problem, the left side is as follows:

$$\begin{aligned} E(Y(t)|\mathcal{F}_s) &= E(B^2(t) - t) \\ &= E((B(t) - B(s) + B(s))^2 - (t - s + s)) \\ &= E((B^2(s) - s) + 2B(s)(B(t) - B(s)) + (B(t) - B(s))^2 - (t - s)) \\ &= Y(s) + 2B(s)E(B(t) - B(s)) + E((B(t) - B(s))^2) - (t - s) \end{aligned}$$

Recall that for a standard Brownian motion process $B(t)$, $E(B(t) - B(s)) = 0$. Hence, $E((B(t) - B(s))^2) = Var(B(t - s)) = t - s$. So we have

$$\begin{aligned} &Y(s) + 2B(s)E(B(t) - B(s)) + E((B(t) - B(s))^2) - (t - s) \\ &= Y(s) + 0 + (t - s) - (t - s) \\ &= Y(s) \end{aligned}$$

thus verifying the Martingale property.

For the expected value, we have:

$$\begin{aligned} E(B^2(t) - t) &= E(B^2(t)) - t \\ &= Var(B(t)) - t \\ &= t - t \\ &= 0 \end{aligned}$$

Note: Suppose $\{Y(t), t \geq 0\}$ is a Martingale process and you continually observe it up to some *stopping time* T. Then we have

$$E(Y(T)) = E(Y(0))$$

This result is known as the *Martingale stopping theorem.*

11. Let T be the first time that the standard Brownian motion hits time $2 - 4t$:

$$T = \min\{t : B(t) = 2 - 4t\}$$

Solution: By the Martingale stopping theorem, $E(B(t)) = E(B(0)) = 0$, but on the left hand side of this equation, we have

$$\begin{aligned} E(B(t)) &= E(2 - 4T) \\ &= 2 - 4E(T) \end{aligned}$$

Solving the equation $2 - 4E(T) = 0$ gives us that $E(T) = \frac{1}{2}$.

12. Let $\{X(t), t \geq 0\}$ be the Brownian motion with drift coefficient $\mu > 0$ and variance parameter σ^2:

$$X(t) = \sigma B(t) + \mu t$$

Further let T be the first time the process reaches value x:

$$\begin{aligned} T &= \min\{t : X(t) = x\} \\ &= \min\left\{t : B(t) = \frac{x - \mu t}{\sigma}\right\} \end{aligned}$$

Use the Martingale stopping theorem to show that

$$E(T) = \tfrac{x}{\mu}$$

Solution: $X(t) = \sigma B(t) + \mu t \Longrightarrow B(t) = \frac{X(t) - \mu t}{\sigma}$. So from this, we have

$$\begin{aligned} T &= \min\{t : X(t) = x\} \\ &= \min\left\{t : B(t) = \frac{x - \mu t}{\sigma}\right\} \end{aligned}$$

By the Martingale stopping theorem, $E(B(T)) = E(B(0)) = 0$. And the left side of the equation equals

$$\begin{aligned} E(B(T)) &= \frac{1}{\sigma}(E(X(T)) - \mu E(T)) \\ &= \frac{1}{\sigma}(x - \mu E(T)) \end{aligned}$$

Solving for $E(T)$:

$$\begin{aligned} \frac{1}{\sigma}(x - \mu E(T)) &= 0 \\ x - \mu E(T) &= 0 \\ E(T) &= \frac{x}{\mu} \end{aligned}$$

13. Let $X(t) = \sigma B(t) + \mu t$ and for given a and b, let p be the probability that $X(t)$ hits a before it hits $-b$.

    (a) Define the stopping time T to be the first time the process hits either a or $-b$. Use this stopping time and the following martingale $\{Y(t) : t \geq 0\}$:

    $$Y(t) = e^{cB(t) - \frac{c^2 t}{2}}$$

    to show that

    $$E\left(\exp\left(\frac{cX(T) - \mu T}{\sigma} - \frac{c^2 T}{2}\right)\right) = 1$$

    Solution: In mathematical notation: $T = \{t : X(t) = a \text{ or } X(t) = -b\}$. Furthermore, rewrite

    $$X(t) = \sigma B(t) + \mu t \Longrightarrow B(t) = \frac{X(t) - \mu t}{\sigma}$$

    Hence,

    $$\begin{aligned} Y(t) &= e^{cB(t) - \frac{c^2 t}{2}} \\ &= e^{\frac{c}{\sigma}(X(t) - \mu t) - \frac{c^2 t}{2}} \end{aligned}$$

    Recall from previous problem that for $Y(t)$, $E(Y(T)) = 1$. Hence it follows that

    $$E(e^{\frac{c}{\sigma}(X(T) - \mu T) - \frac{c^2 T}{2}}) = 1$$

    (b) Let $c = -\frac{2\mu}{\sigma}$ and show that

    $$E(e^{-\frac{2\mu X(T)}{\sigma^2}}) = 1$$

Solution: Plugging in this value of c:

$$\begin{aligned}
&E\left(e^{\frac{c}{\sigma}(X(T)-\mu T)-\frac{c^2 T}{2}}\right)\\
&=E\left(e^{-\frac{2\mu}{\sigma^2}(X(T)-\mu T)-\frac{4\mu^2 T}{\sigma^2 * 2}}\right)\\
&=E\left(e^{-\frac{2\mu}{\sigma^2}(X(T)-\mu T)-\frac{2\mu^2 T}{\sigma^2}}\right)\\
&=E\left(e^{-\frac{2\mu}{\sigma^2}(X(T)-\mu T+\mu T)}\right)\\
&=E\left(e^{-\frac{2\mu}{\sigma^2}(X(T)}\right)
\end{aligned}$$

(c) Use the previous part and the definition of T to find p.
Solution: Recall that p denotes the probability of hitting a before $-b$. Then

$$X(T)=\begin{cases} a \text{ w.p. } p \\ -b \text{ w.p. } 1-p \end{cases}$$

Following the definition of expectation:

$$\begin{aligned}
E\left(e^{-\frac{2\mu}{\sigma^2}(X(T)}\right) &= p * e^{-\frac{2\mu}{\sigma^2}a}+(1-p) * e^{-\frac{2\mu}{\sigma^2}(-b)}\\
&= e^{\frac{2\mu}{\sigma^2}b}+p\left(e^{-\frac{2\mu}{\sigma^2}a}-e^{-\frac{2\mu}{\sigma^2}b}\right)
\end{aligned}$$

From the previous parts, we know that $E\left(e^{-\frac{2\mu}{\sigma^2}(X(T)}\right)=1$. So we can solve for p to get

$$p=\frac{1-e^{\frac{2\mu}{\sigma^2}b}}{e^{-\frac{2\mu}{\sigma^2}a}-e^{\frac{2\mu}{\sigma^2}b}}$$

14. Let $X(t)=\sigma B(t)+\mu t$ and define T to be the first time the process $\{X(t), t \geq 0\}$ hits either a or $-b$, where $a, b>0$. Use the Martingale stopping theorem and the last part of the previous exercise to find $E(T)$.
Solution: Since $X(t)=\sigma B(t)+\mu t$, we have

$$\begin{gathered}
X(T)=\sigma B(T)+\mu T\\
E(X(T))=\sigma E(B(T))+\mu E(T)
\end{gathered}$$

For standard Brownian motion process $B(t)$, $E(B(t))=0$. And from the previous part we have:

$$E(X(T))=p * a-(1-p) * b$$

Hence, solving for $E(T)$, we have

$$E(T)=\frac{p * a-(1-p) * b}{\mu}$$

15. Let $\{X(t), t \geq 0\}$ be Brownian motion with drift coefficient $\mu > 0$ and variance parameter σ^2:

$$X(t) = \mu t + \sigma B(t)$$

Let $x > 0$ and define the stopping time T to be the same as in a previous problem:

$$T = \min\{t : X(t) = x\}$$

Use the following martingale $\{Y(t), t \geq 0\}$:

$$Y(t) = B(t)^2 - t$$

and the fact from a previous exercise:

$$E(T) = \tfrac{x}{\mu}$$

to show that

$$\mathrm{Var}(T) = \tfrac{x\sigma^2}{\mu^3}$$

Solution: Recall that

$$E(B^2(T) - T) = 0$$

and again, let us rewrite the equation $X(t) = \mu t + \sigma B(t)$ to

$$B(t) = \tfrac{X(t) - \mu t}{\sigma}$$

Substituting:

$$\begin{aligned}
0 &= E\left(\left(\frac{X(T) - \mu T}{\sigma}\right)^2 - T\right) \\
&= E\left(\frac{1}{\sigma^2}(X(T) - \mu T)^2 - T\right) \\
&= E\left(\frac{1}{\sigma^2}(x - \mu T)^2 - T\right) \\
&= E\left(\frac{1}{\sigma^2}\left(x^2 - 2x\mu T + \mu^2 T^2\right) - T\right) \\
&= \frac{x^2}{\sigma^2} - \frac{2x\mu}{\sigma^2}E(T) + \frac{\mu^2}{\sigma^2}E(T^2) - E(T) \\
&= \frac{x^2}{\sigma^2} - \left(\frac{2x\mu}{\sigma^2} + 1\right)E(T) + \frac{\mu^2}{\sigma^2}E(T^2)
\end{aligned}$$

Recall $E(T) = \frac{x}{\mu}$.

Note: In order to solve the following problem, we shall introduce some background material.

Let f be a function that is continuous in the interval $[a, b]$. Then we define the *stochastic integral* as follows:

$$\int_a^b f(t)dB(t) = \lim_{n\to\infty} \lim_{\Delta t_i \to 0} \sum_{i=1}^{n} f(t_{i-1})(B(t_i) - B(t_{i-1}))$$

where $a = t_0 < t_1 < t_2 < \cdots < t_n = b$ is a partition of the interval $[a, b]$ and $\Delta t_i = t_i - t_{i-1}$ and we take the limit over all Δt_i.

We can try simplifying the sum on the right. First consider the smaller case where $n = 3$. For sake of ease in notation, let $t_i = i$ for all i:

$$\begin{aligned}
&\sum_{i=1}^{3} f(i-1)(B(i) - B(i-1)) \\
&= f(0)(B(1) - B(0)) + f(1)(B(2) - B(1)) + f(2)(B(3) - B(2)) \\
&= -f(0)B(0) + B(1)(f(0) - f(1)) + B(2)(f(1) - f(2)) + f(2)B(3) \\
&= f(2)B(3) - f(0)B(0) - \sum_{i=1}^{2} B(i)(f(i) - f(i-1))
\end{aligned}$$

So for generic n, we simplify the sum on the right side as follows:

$$\begin{aligned}
&\sum_{i=1}^{n} f(t_{i-1})(B(t_i) - B(t_{i-1})) \\
&= f(t_{n-1})B(t_n) - f(t_0)B(t_0) - \sum_{i=1}^{n-1} B(t_i)(f(t_i) - f(t_{i-1}))
\end{aligned}$$

and so, by taking the limit:

$$\int_a^b f(t)dB(t) = f(b)B(b) - f(a)B(a) - \int_a^b B(t)df(t)$$

As a sidenote, many may recognize this as the integration by parts formulation:

$$\begin{aligned}
\int_a^b fdv &= f * v\Big|_a^b - \int_a^b vdf \\
&= f(b)v(b) - f(a)v(a) - \int_a^b vdf
\end{aligned}$$

Taking the expected value of both sides:

$$\begin{aligned}
E\left(\int_a^b f(t)dB(t)\right) &= E(f(b)B(b) - f(a)B(a) - \int_a^b B(t)df(t)) \\
&= f(b)E(B(b)) - f(a)E(B(a)) - \int_a^b E(B(t))df(t)) \\
&= f(b) * 0 - f(a) * 0 - \int_a^b 0)df(t)) \\
&= 0
\end{aligned}$$

To compute the variance, take the variance of the summation form first:

$$\begin{aligned}
Var\left(\sum_{i=1}^{n} f(t_{i-1})(B(t_i) - B(t_{i-1}))\right) &= \sum_{i=1}^{n} Var(f(t_{i-1})(B(t_i) - B(t_{i-1}))) \\
&= \sum_{i=1}^{n} f^2(t_{i-1})Var((B(t_i) - B(t_{i-1})) \\
&= \sum_{i=1}^{n} f^2(t_{i-1})(t_i - t_{i-1})
\end{aligned}$$

Taking limits as above, we get that

$$Var\left(\int_a^b f(t)dB(t)\right) = \int_a^b f^2(t)dt$$

16. Compute the mean and variance of

    (a) $\int_0^1 tdB(t)$
    Solution: By the above Note, the mean of this quantity is zero, and the variance is:

    $$\begin{aligned}
    Var\left(\int_0^1 tdB(t)\right) &= \int_0^1 t^2 dt \\
    &= \frac{1}{3}
    \end{aligned}$$

    (b) $\int_0^1 t^2 dB(t)$
    Solution: Again, the mean of this quantity is zero. The variance is:

    $$\begin{aligned}
    Var\left(\int_0^1 t^2 dB(t)\right) &= \int_0^1 t^4 dt \\
    &= \frac{1}{5}
    \end{aligned}$$

17. Let $Y(t) = tB\left(\frac{1}{t}\right), t > 0$ and $Y(0) = 0$

    (a) What is the distribution of $Y(t)$?
    Solution: $Y(t)$ is normal with mean 0:

    $$\begin{aligned}
    E(Y(t)) &= tE\left(B\left(\frac{1}{t}\right)\right) \\
    &= 0
    \end{aligned}$$

    because $B(t)$ is the standard Brownian motion process.

    $Y(t)$ is normal with mean 0 and variance t:

    $$\begin{aligned}
    Var\left(tB\left(\frac{1}{t}\right)\right) &= t^2 Var\left(B\left(\frac{1}{t}\right)\right) \\
    &= t^2 * \frac{1}{t} \\
    &= t
    \end{aligned}$$

(b) Find $\mathrm{Cov}(Y(s), Y(t))$.
Solution:

$$\begin{aligned}\mathrm{Cov}(Y(s), Y(t)) &= E(Y(s)Y(t)) \\ &= E\left(sB\left(\frac{1}{s}\right) * tB\left(\frac{1}{t}\right)\right) \\ &= stE\left(B\left(\frac{1}{s} * \frac{1}{t}\right)\right)\end{aligned}$$

Let $s \leq t$. Then

$$E\left(B\left(\frac{1}{s} * \frac{1}{t}\right)\right) = \frac{1}{t}$$

Hence

$$\begin{aligned}&stE\left(B\left(\frac{1}{s} * \frac{1}{t}\right)\right) \\ &= st * \frac{1}{t} \\ &= s\end{aligned}$$

(c) Argue that $\{Y(t), t \geq 0\}$ is a standard Brownian motion process.
Solution: We have the same distribution and the same covariance function. Hence, it must be the same process.

18. Let $Y(t) = \frac{1}{a}B(a^2t)$ for $a > 0$. Argue that $\{Y(t)\}$ is a standard Brownian motion process.
Solution: We know that $Y(t)$ is Gaussian distributed. We shall then compute the mean:

$$\begin{aligned}E(Y(t)) &= \frac{1}{a}E(B(a^2t)) \\ &= \frac{1}{a} * 0 \\ &= 0\end{aligned}$$

and the variance.

$$\begin{aligned}Var(Y(t)) &= \frac{1}{a^2}Var(B(a^2t)) \\ &= \frac{1}{a^2} * a^2t \\ &= t\end{aligned}$$

Now compute the covariance function:

$$\begin{aligned}\mathrm{Cov}(Y(s), Y(t)) &= E\left(\frac{1}{a^2}B(a^2s)B(a^2t)\right) \\ &= \frac{1}{a^2}E\left(B(a^2s)B(a^2t)\right) \\ &= \frac{1}{a^2} * a^2s \\ &= s\end{aligned}$$

Therefore, $Y(t)$ must be the same process as $B(t)$.

19. For $s < t$, argue that $B(s) - \frac{s}{t}B(t)$ and $B(t)$ are independent.
Solution: We shall look at the covariance between these two:

$$\begin{aligned}
\text{Cov}(B(s) - \frac{s}{t}B(t), B(t)) &= \text{Cov}(B(s), B(t)) - \frac{s}{t}Var(B(t)) \\
&= s - \frac{s}{t} * t \\
&= 0
\end{aligned}$$

Hence, it follows that the two are independent.

20. Let $\{Z(t), t \geq 0\}$ denote a Brownian bridge process. Show that if

$$Y(t) = (t+1)Z\left(\frac{t}{t+1}\right)$$

then $\{Y(t), t \geq 0\}$ is a standard Brownian motion process.
Solution: $\{Z(t), t \geq 0\}$ is a Brownian bridge process, so:

$$Z(t) = B(t) - tB(1)$$

which is Gaussian distributed.

Again, we will check the mean and the variance of $Y(t)$.

$$\begin{aligned}
E(Y(t)) &= (t+1)E\left[Z\left(\frac{t}{t+1}\right)\right] \\
&= (t+1)E\left[Z\left(\frac{t}{t+1}\right) - tZ\left(\frac{1}{2}\right)\right] \\
&= (t+1) * 0 \\
&= 0
\end{aligned}$$

and

$$\begin{aligned}
Var(Y(t)) &= (t+1)^2 Var\left[Z\left(\frac{t}{t+1}\right)\right] \\
&= (t+1)^2\left(\frac{t}{t+1}\right)\left(1 - \frac{t}{t+1}\right) \\
&= (t+1)^2\left(\frac{t}{t+1}\right)\left(\frac{1}{t+1}\right) \\
&= t
\end{aligned}$$

Hence $Y(t)$ is Gaussian distributed with mean 0 and variance t. Now we must check the covariance:

$$\begin{aligned}
\mathrm{Cov}(Y(s), Y(t)) &= \mathrm{Cov}\left[(s+1)Z\left(\frac{s}{s+1}\right), (t+1)Z\left(\frac{t}{t+1}\right)\right] \\
&= (s+1)(t+1)\mathrm{Cov}\left[Z\left(\frac{s}{s+1}\right), Z\left(\frac{t}{t+1}\right)\right] \\
&= (s+1)(t+1)\left(\frac{s}{s+1}\right)\left(1-\frac{t}{t+1}\right) \\
&= (s+1)(t+1)\left(\frac{s}{s+1}\right)\left(\frac{1}{t+1}\right) \\
&= s
\end{aligned}$$

Wide Sense Stationarity

First, we formally define a *random process* $X(t)$ to be a random variable X that is a function of time t. Random processes can be used in various applications, including modeling a random signal or random noise process at a receiver.

Let $X(t)$ be characterized by the PDF $f_X(x,t)$. Then we have the following properties:

1. The *mean* $\mu_X(t)$ is a function of time:

$$\mu_{X(t)} = E(X(t)) = \int_{-\infty}^{\infty} x f_x(x,t) dx$$

2. The *autocorrelation* function is the correlation between the random variables at the two times t_1 and t_2:

$$\begin{aligned} R_{XX}(t_1,t_2) &= E(X(t_1), X(t_2)) \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1 x_2 f(x_1,x_2,t_1,t_2) dx_1 dx_2 \end{aligned}$$

   where $f(x_1,x_2,t_1,t_2)$ denotes the joint probability density for t_1 and t_2.

$\{X(t), t \geq 0\}$ is a *stationary process* if for all $s, t, t_1, t_2, \cdots t_n$, the random variables $X(t_1), X(t_2), \cdots X(t_n)$ have the same joint distribution as that of $X(t_1+s), X(t_2+s), \cdots X(t_n+s)$.

We shall consider a couple of examples:

1. An ergodic continuous-time Markov chain $\{X(t), t \geq 0\}$ with $P(X(0) = j) = P_j, j \geq 0$ is a stationary process because it is a Markov chain with initial state chosen according to the limiting probabilities. So it can be regarded as an ergodic Markov chain that we start observing at some very large time t. Then a continuation of the process at time s is just a continuation of the chain starting at $t+s$ so it follows that the limiting probabilities are the same for all s.

2. $\{X(t), t \geq 0\}$ (where $X(t) = N(t+L) - N(t), t \geq 0$, L a fixed constant, and $N(t)$ a Poisson process with rate λ) is a stationary process. This is because $X(t)$ is the number of events of the Poisson process that occur between t and $t+L$, and the increments of a Poisson process are stationary.

Now a *Wide Sense Stationary* (WSS) process (also called a *Weakly Stationary* process) is a random process that has the following special properties:

1. The mean $\mu_{X(t)}$ is a constant. In other words, it is stationary in the mean.

2. Let $t_1 = t, t_2 = t + \tau$. Then

$$\begin{aligned} E(X(t)X(t+\tau)) &= R_{XX}(t, t+\tau) \\ &= R_{XX}(\tau) \end{aligned}$$

   In other words, it is stationary in the sense that only the difference in time matters, not what time it happened.

If we set $\tau = 0$, we get

$$\begin{aligned} E(X(t)X(t)) &= E(X^2(t)) \\ &= R_{XX}(0) \end{aligned}$$

Further note that $E(X^2(t))$ is the average power in the random process $X(t)$. It then follows that $R_{XX}(0)$ denotes the power in the random process.

Now consider the example of wireless communication. Let the process be defined as follows:

$$X(t) = \alpha \cos(2\pi f_c t + \theta)$$

Let θ be random and uniformly distributed over $[-\pi, \pi]$:

$$f_\Theta(\theta) = \begin{cases} \frac{1}{2\pi} \text{ if } -\pi \leq \theta \leq \pi \\ 0 \text{ else} \end{cases}$$

and let α and f_c be constants. We wish to determine if $X(t)$ is WSS.

We shall check the two properties that constitute a WSS process.

1.

$$\begin{aligned} E(X(t)) &= E(\alpha \cos(2\pi f_c t + \theta)) \\ &= \int_{-\pi}^{\pi} \alpha \cos(2\pi f_c t + \theta) f_\Theta(\theta) d\theta \\ &= \frac{\alpha}{2\pi} \int_{-\pi}^{\pi} \alpha \cos(2\pi f_c t + \theta) d\theta \\ &= \frac{\alpha}{2\pi} (\sin(2\pi f_c t + \pi) - \sin(2\pi f_c t - \pi)) d\theta \\ &= \frac{\alpha}{2\pi} * 0 \\ &= 0 \end{aligned}$$

which is constant. Hence the process is stationary in the mean.

2.

$$
\begin{aligned}
E(X(t)X(t+\tau)) &= E(\alpha\cos(2\pi f_c t+\theta)\alpha\cos(2\pi f_c(t+\tau)+\theta)) \\
&= \int_{-\pi}^{\pi} \alpha\cos(2\pi f_c t+\theta)\alpha\cos(2\pi f_c(t+\tau)+\theta) f_\Theta(\theta) d\theta \\
&= \frac{\alpha^2}{2\pi}\int_{-\pi}^{\pi} \cos(2\pi f_c t+\theta)\cos(2\pi f_c(t+\tau)+\theta) d\theta \\
&= \frac{\alpha^2}{2\pi}\int_{-\pi}^{\pi} \frac{1}{2}(\cos(2\pi f_c\tau)+\cos(4\pi f_c(t+2\pi f_c\tau)+2\theta)) d\theta \\
&= \frac{\alpha^2}{4\pi}\int_{-\pi}^{\pi} (\cos(2\pi f_c\tau)+\cos(4\pi f_c(t+2\pi f_c\tau)+2\theta)) d\theta \\
&= \frac{\alpha^2}{4\pi}\left(\cos(2\pi f_c\tau)2\pi + \frac{1}{2}\sin(4\pi f_c(t+2\pi f_c\tau)+2\theta)\Big|_{-\pi}^{\pi}\right) \\
&= \frac{\alpha^2}{4\pi}\cos(2\pi f_c\tau)2\pi \\
&= \frac{\alpha^2}{2\pi}\cos(2\pi f_c\tau)
\end{aligned}
$$

which depends only on τ and not on t. Hence the process is stationary in its autocorrelation. In conclusion, the process is WSS.

Furthermore,

$$R_{XX}(0) = \tfrac{\alpha^2}{2}\cos(2\pi f_c * 0) = \tfrac{\alpha^2}{2}$$

is the power of the process.

We define the *power spectral density* (PSD) of a WSS process $X(t)$ to be

$$S_{XX}(f) = \int_{-\infty}^{\infty} R_{XX}(\tau)e^{-j2\pi f\tau}d\tau$$

Note that this is the Fourier transform of the autocorrelation function.

Therefore the autocorrelation $R_{XX}(\tau)$ is the inverse Fourier transform of the PSD.

$$R_{XX}(\tau) = \int_{-\infty}^{\infty} S_{XX}(f)e^{j2\pi f\tau}df$$

Some properties of the PSD are as follows:

1. $S_{XX}(f) \geq 0$ and is a real quantity.
2. $S_{XX}(f)$ is symmetric about $f = 0$: $= S_{XX}(-f)$.

Setting $\tau = 0$ above, we get

$$
\begin{aligned}
R_{XX}(0) &= \int_{-\infty}^{\infty} S_{XX}(f)e^{j2\pi f*0}df \\
&= \int_{-\infty}^{\infty} S_{XX}(f)df
\end{aligned}
$$

hence, the area under the PSD from $-\infty$ to ∞ is the power.

We will now consider an application of the power spectral density.

Consider a wireless signal $X(t)$ with autocorrelation:

$$R_{XX}(\tau) = \tfrac{1}{2a}e^{-a|\tau|}$$

and let a be 5kHz.

What is the power? What is the PSD? Furthermore, from the PSD, calculate the bandwidth required which contains 90% of the signal power.

The power is given to be

$$\begin{aligned}
E(X^2(t)) &= R_{XX}(0) \\
&= \frac{1}{2a}e^{-a|0|} \\
&= \frac{1}{2a}
\end{aligned}$$

The PSD is given to be

$$\begin{aligned}
S_{XX}(f) &= \int_{-\infty}^{\infty} R_{XX}(\tau)e^{-j2\pi f\tau}d\tau \\
&= \int_{-\infty}^{\infty} \frac{1}{2a}e^{-a|\tau|}e^{-j2\pi f\tau}d\tau \\
&= \frac{1}{2a}(\int_{-\infty}^{0} e^{a\tau}e^{-j2\pi f\tau}d\tau + \int_{0}^{\infty} e^{-a\tau}e^{-j2\pi f\tau}d\tau) \\
&= \frac{1}{2a}(\int_{-\infty}^{0} e^{(a-j2\pi f)\tau}d\tau + \int_{0}^{\infty} e^{-a\tau}e^{-(a+j2\pi f)\tau}d\tau) \\
&= \frac{1}{2a}\left(\frac{1}{(a-j2\pi f)}e^{(a-j2\pi f)\tau}\Big|_{-\infty}^{0} - \frac{1}{(a+j2\pi f)}e^{-a\tau}e^{-(a+j2\pi f)\tau}\Big|_{0}^{\infty}\right) \\
&= \frac{1}{2a}\left(\frac{1}{(a-j2\pi f)} + \frac{1}{(a+j2\pi f)}\right) \\
&= \frac{1}{2a}\left(\frac{2a}{(a^2+(2\pi f)^2)}\right) \\
&= \frac{1}{a^2+(2\pi f)^2} \\
&= S_{XX}(f)
\end{aligned}$$

Note that at $f=0$, $S_{XX}(f) = \frac{1}{a^2}$ and tends to 0 as f goes to ∞ or $-\infty$.

Suppose the bandwidth that under which 90% of the power is in is $[-W, W]$. Then according to formula, the energy of $X(t)$ in the bandwidth is

$$\int_{-W}^{W} S_{XX}(f)df$$

We desire this to be $0.9 * \frac{1}{2a}$:

$$
\begin{aligned}
0.9 * \frac{1}{2a} &= \int_{-W}^{W} \frac{1}{a^2 + (2\pi f)^2} df \\
&= \frac{1}{a^2} \int_{-W}^{W} \frac{1}{1 + \left(\frac{2\pi}{a} f\right)^2} df
\end{aligned}
$$

Recall that

$$\int \frac{1}{x^2+1} dx = \tan^{-1}(x)$$

So using the substitution $x = \frac{2\pi}{a} f$ and $dx = \frac{2\pi}{a} df$:

$$
\begin{aligned}
&\frac{1}{a^2} \int_{-W}^{W} \frac{1}{1 + \left(\frac{2\pi}{a} f\right)^2} df \\
&= \frac{1}{a^2} \frac{a}{2\pi} \int_{-\frac{2\pi W}{a}}^{\frac{2\pi W}{a}} \frac{1}{1 + x^2} dx \\
&= \frac{1}{2\pi a} \tan^{-1}(x) \Big|_{-\frac{2\pi W}{a}}^{\frac{2\pi W}{a}} \\
&= \frac{1}{2\pi a} * 2 \tan^{-1} \left(\frac{2\pi W}{a} \right)
\end{aligned}
$$

Hence $\tan^{-1}(\frac{2\pi W}{a}) = \frac{0.9\pi}{2}$. So $W = 1.005a$, and the *passband width* is $2.010a$.

We will now look at the transmission of a random process through an LTI system. Consider $X(t)$, a WSS process, and $h(t)$ the impulse response of the LTI system. Let $Y(t)$ be the output of this process through the system.

$$
\begin{aligned}
Y(t) &= X(t) * h(t) \\
\implies Y(t) &= \int_{-\infty}^{\infty} X(t - \alpha) h(\alpha) d\alpha
\end{aligned}
$$

Let us examine some properties of $Y(t)$. We know that it is a random process; now we shall check if it is WSS.

1. $$
\begin{aligned}
E(Y(t)) &= E(Y(t)) \\
&= E\left(\int_{-\infty}^{\infty} X(t - \alpha) h(\alpha) d\alpha \right) \\
&= \int_{-\infty}^{\infty} E(X(t - \alpha)) h(\alpha) d\alpha \\
&= \int_{-\infty}^{\infty} \mu_X h(\alpha) d\alpha
\end{aligned}
$$

   where the last line comes from the fact that $X(t)$ is WSS, so $E(X(t)) = \mu_X$ is constant.

But note that the integral above is independent of time. Hence, $Y(t)$ is stationary with respect to its mean.

$$\mu_Y = \mu_X \int_{-\infty}^{\infty} h(\alpha) d\alpha$$

2. Since $Y(t) = \int_{-\infty}^{\infty} X(t-\alpha)h(\alpha)d\alpha$, we have $Y(t+\tau) = \int_{-\infty}^{\infty} X(t+\tau-\beta)h(\beta)d\beta$.

   So we get

$$\begin{aligned}
&E(Y(t)Y(t+\tau)) \\
&= E\left(\left(\int_{-\infty}^{\infty} X(t-\alpha)h(\alpha)d\alpha\right)\left(\int_{-\infty}^{\infty} X(t+\tau-\beta)h(\beta)d\beta\right)\right) \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} E(X(t-\alpha)X(t+\tau-\beta))h(\alpha)h(\beta)d\alpha d\beta
\end{aligned}$$

   Now note that $E(X(t-\alpha)X(t+\tau-\beta))$ is the expression for the autocorrelation of X at $t_1 = t-\alpha$ and $t_2 = t+\tau-\beta$:

$$R_{XX}(\tau-\beta+\alpha) = E(X(t-\alpha)X(t+\tau-\beta))$$

   Continuing our calculations:

$$\begin{aligned}
&\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} E(X(t-\alpha)X(t+\tau-\beta))h(\alpha)h(\beta)d\alpha d\beta \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} R_{XX}(\tau-\beta+\alpha)h(\alpha)h(\beta)d\alpha d\beta
\end{aligned}$$

   which is an expression that will depend only on the time shift τ, not on exactly at which time it occurs.

   Hence $Y(t)$ is stationary in autocorrelation.

$$R_{YY}(\tau) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} R_{XX}(\tau-\beta+\alpha)h(\alpha)h(\beta)d\alpha d\beta$$

In conclusion, $Y(t)$ is a WSS process.

Let us now substitute $\tilde{\alpha} = -\alpha$ and $d\tilde{\alpha} = -d\alpha$:

$$\begin{aligned}
R_{YY}(\tau) &= -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} R_{XX}(\tau-\beta-\tilde{\alpha})h(-\tilde{\alpha})h(\beta)d\tilde{\alpha}d\beta \\
&= \int_{\infty}^{-\infty}\int_{-\infty}^{\infty} R_{XX}(\tau-\beta-\tilde{\alpha})h(-\tilde{\alpha})h(\beta)d\tilde{\alpha}d\beta \\
&= \int_{\infty}^{-\infty}\int_{-\infty}^{\infty} R_{XX}(\tau-\beta-\tilde{\alpha})\tilde{h}(\alpha)h(\beta)d\tilde{\alpha}d\beta
\end{aligned}$$

Observe that this is basically a convolution:

$$R_{YY}(\tau) = R_{XX}(\tau) * h(\tau) * \tilde{h}(\tau)$$

Let us see what this gives us in the frequency domain. First we have

$$\begin{aligned}
\tilde{h}(\tau) &\longleftrightarrow \tilde{H}(f) \\
&= \int_{-\infty}^{\infty} h(-\tau) e^{-j2\pi f\tau} d\tau \\
&= \int_{-\infty}^{\infty} h(\tilde{\tau}) e^{j2\pi f\tilde{\tau}} d\tilde{\tau}
\end{aligned}$$

and

$$\begin{aligned}
\tilde{H}^*(f) &= \left(\int_{-\infty}^{\infty} h(\tilde{\tau}) e^{j2\pi f\tilde{\tau}} d\tilde{\tau} \right)^* \\
&= \int_{-\infty}^{\infty} h^*(\tilde{\tau}) e^{-j2\pi f\tilde{\tau}} d\tilde{\tau} \\
&= \int_{-\infty}^{\infty} h(\tilde{\tau}) e^{-j2\pi f\tilde{\tau}} d\tilde{\tau} \\
&= H(f)
\end{aligned}$$

where $H(f)$ is the Fourier transform of $h(\tau)$.

So we have

$$h(-\tau) = \tilde{h}(\tau) \longleftrightarrow \tilde{H}(f) = H^*(f)$$

because $\tilde{H}^*(f) = H(f)$, so $\tilde{H}(f) = H^*(f)$

Taking the Fourier Transform of the whole equation $R_{YY}(\tau) = R_{XX}(\tau) * h(\tau) * \tilde{h}(\tau)$:

$$\begin{aligned}
S_{YY}(f) &= S_{XX}(f) * H(f) * \tilde{H}(f) \\
&= S_{XX}(f) * H(f) * H^*(f) \\
&= S_{XX}(f)|H(f)|^2
\end{aligned}$$

We will now introduce a few different kinds of special random processes.

1. **Gaussian Random Process**

   A random process $X(t)$ is a Gaussian random process if $X(t_1), X(t_2), \cdots X(t_n)$ are jointly Gaussian. That is, the multivariate density $f_{X(t_1),\cdots,X(t_n)}(x_1, \cdots, x_n)$ is Gaussian for all n and for all time points $t_1, t_2, \cdots t_n$.

   In particular, for $n = 1$, the PDF of $X(t)$ at any time t must be Gaussian:

   $$f_{X(t)}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

   where

   $$\begin{aligned}
   \mu &= E(X(t)) \\
   \sigma^2 &= E((X(t) - \mu)^2)
   \end{aligned}$$

Now consider a WSS Gaussian Process. This means that

$$\begin{aligned} E(X(t)) &= \mu_X, \text{ constant} \\ E(X(t)X(t+\tau)) &= R_{XX}(\tau) \end{aligned}$$

and when $\tau = 0$

$$\begin{aligned} E(X^2(t)) &= R_{XX}(0) \\ \sigma^2 &= E(X^2(t)) - {\mu_X}^2 \\ &= R_{XX}(0) - {\mu_X}^2 \end{aligned}$$

We also have the following probability density of $X(t)$ at any time t for a Gaussian random process, which is also WSS.

$$\begin{aligned} f_{X(t)}(x) &= \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} \\ &= \frac{1}{\sqrt{2\pi(R_{XX}(0) - {\mu_X}^2)}} e^{\frac{-(x-\mu)^2}{2(R_{XX}(0)-{\mu_X}^2)}} \end{aligned}$$

The standard Brownian motion process $B(t)$ is a Gaussian process because each $X(t_1), X(t_2), \cdots, X(t_n)$ can be expressed as a linear combination of the independent Gaussian random variables $X(t_1), X(t_2) - X(t_1), X(t_3) - X(t_2), \cdots, X(t_n)$.

2. **White Noise/ White Process**

A random process $X(t)$ is white if it is WSS and

$$R_{XX}(\tau) = \tfrac{\eta}{2}\delta(\tau)$$

where $\delta(\tau)$ represents that there is an impulse at $\tau = 0$.

Observe that this means $R_{XX}(\tau) = 0$ if $\tau \neq 0$.

This, in turn, implies that $E(X(t)X(t+\tau)) = 0$ if $\tau \neq 0$. So $X(t)$ and $X(t+\tau)$ are uncorrelated for any $\tau \neq 0$.

Let us consider the PSD (power spectral density) of white noise:

$$\begin{aligned} S_{XX}(f) &= \mathcal{F}\{R_{XX}(\tau)\} \\ &= \mathcal{F}\{\frac{\eta}{2}\delta(\tau)\} \\ &= \frac{\eta}{2} \text{ for all } -\infty < f < \infty \end{aligned}$$

3. **White Gaussian Noise**

A random process which is both Gaussian and white is a White Gaussian process.

We will consider an example in the context of communication systems. Suppose any transmitted signal $x(t)$ gets some Gaussian white noise $n(t)$ added to it, and $y(t)$ is the received signal. That is,

$$y(t) = x(t) + n(t)$$

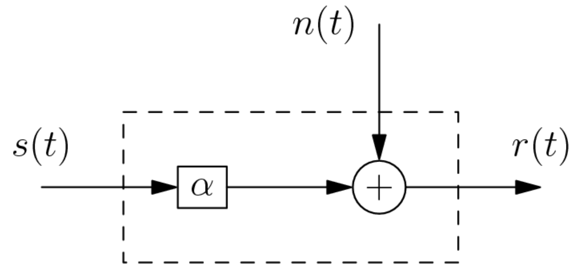


diagram above obtained from Google Images. In our case, $s(t) = x(t)$ and $r(t) = y(t)$.

Such a configuration is called an "Additive White Gaussian Noise" (AWGN) channel. This is a popular model, typically used to model a wireline communication channel. If we were to model a wireless communication channel instead, we would add an additional *fading coefficient* to the transmitted signal and the noise to obtain the received signal.

Now recall that when we transmit a random process $X(t)$ through a LTI system to obtain another random process $Y(t)$, we have

$$S_{YY}(f) = S_{XX}(f)|H(f)|^2$$

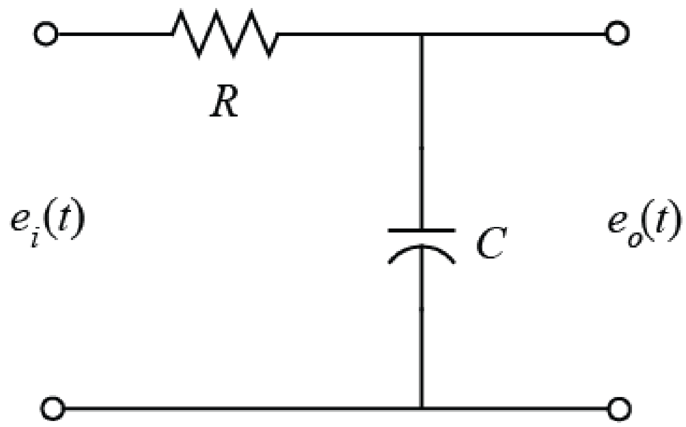
Further recall that is $X_1, \cdots, X_n$ are Gaussian random variabler, then

$$X = a_1 X_1 + \cdots + a_n X_n$$

is also Gaussian.

Suppose we input a Gaussian random process into a LTI system. By the property described above, if $Y(t)$ is Gaussian.

Now suppose we have white noise. We shall continue our analysis in the form of the following example, where some zero-mean white Gaussian noise $X(t)$ is passed through an RC low pass filter (LPF). We desire to find the probability density of $Y(t)$, $f_{Y(t)}(y)$.



circuit diagram from Google Images. In our case, $e_i(t) = X(t)$ and $e_o(t) = Y(t)$.

First, we know that $Y(t)$ is Gaussian, as seen above. So all we need to determine are the mean and the variance.

We have that $X(t)$ is WSS and $R_{XX}(\tau) = \frac{\eta}{2}\delta(\tau)$. And in the frequency domain, we get:

$$\begin{aligned} Y(\omega) &= X(\omega)\frac{\frac{1}{j\omega C}}{R + \frac{1}{j\omega C}} \\ &= X(\omega)\frac{1}{1 + j\omega RC} \end{aligned}$$

Hence,

$$\begin{aligned} H(\omega) &= \frac{Y(\omega)}{X(\omega)} \\ &= \frac{1}{1 + j\omega RC} \end{aligned}$$

or

$$H(f) = \frac{1}{1 + j2\pi fRC}$$

So we can see that

$$\begin{aligned} S_{YY}(f) &= S_{XX}(f)|H(f)|^2 \\ &= \frac{\eta}{2}\frac{1}{1 + (2\pi fRC)^2} \end{aligned}$$

And so

$$\begin{aligned} E(|Y(t)|^2) &= \int_{-\infty}^{\infty} S_{YY}(f)df \\ &= \int_{-\infty}^{\infty} \frac{\eta}{2}\frac{1}{1 + (2\pi fRC)^2}df \end{aligned}$$

and once we use the substitution $u = 2\pi RCf$ and $du = 2\pi RCdf$, we get:

$$
\begin{aligned}
&\frac{\eta}{2}\int_{-\infty}^{\infty}\frac{1}{1+(2\pi fRC)^2}df \\
&= \frac{\eta}{2} * \frac{1}{2\pi RC}\int_{-\infty}^{\infty}\frac{1}{1+u^2}du \\
&= \frac{\eta}{4\pi RC}\tan^{-1}(u)\Big|_{-\infty}^{\infty} \\
&= \frac{\eta}{4\pi RC} * \pi \\
&= \frac{\eta}{4RC}
\end{aligned}
$$

So we have

$$R_{YY}(0) = E\left(|Y(t)|^2\right) = \tfrac{\eta}{4RC}$$

From the above, we can finally determine the mean and the variance of $Y(t)$.

$$
\begin{aligned}
\mu_Y &= \left(\int_{-\infty}^{\infty} h(t)dt\right)\mu_X \\
&= 0
\end{aligned}
$$

since $\mu_X = 0$ given. And for the variance,

$$
\begin{aligned}
\sigma^2 &= E(Y^2(t)) - 0^2 \\
&= \frac{\eta}{4RC}
\end{aligned}
$$

So $Y(t)$ is Gaussian with mean 0 and variance $\frac{\eta}{4RC}$.

$$
\begin{aligned}
f_{Y(t)}(y) &= \frac{1}{\sqrt{2\pi\sigma_Y{}^2}}e^{\frac{-(y-\mu_Y)^2}{2\sigma_Y{}^2}} \\
&= \frac{1}{\sqrt{2\pi(\frac{\eta}{4RC})}}e^{\frac{-y^2}{2\frac{\eta}{4RC}}} \\
&= \sqrt{\frac{2RC}{\pi\eta}}e^{\frac{-2RCy^2}{\eta}}
\end{aligned}
$$

Note that $Y(t)$ is not a white process, as the power spectral density is no longer constant!

Stationary Process Problems

1. Let $X(t) = N(t+1) - N(t)$ where $\{N(t), t \geq 0\}$ is a Poisson process with rate λ. Compute

$$\text{Cov}(X(t), X(t+s))$$

Solution: Plug in what we are given and algebraically solve:

$$
\begin{aligned}
&\text{Cov}(X(t), X(t+s)) \\
&\quad = \text{Cov}(N(t+1) - N(t), N(t+s+1) - N(t+s)) \\
&\quad = \text{Cov}(N(t+1), N(t+s+1) - N(t+s)) - \text{Cov}(N(t), N(t+s+1) - N(t+s))
\end{aligned}
$$

and note that $\text{Cov}(N(t), N(t+s+1) - N(t+s)) = 0$ because $N(t)$ and $N(t+s+1) - N(t+s)$ are independent random variables.

So we have

$$
\begin{aligned}
&\text{Cov}(N(t+1), N(t+s+1) - N(t+s)) - \text{Cov}(N(t), N(t+s+1) - N(t+s)) \\
&= \text{Cov}(N(t+1), N(t+s+1) - N(t+s)) \\
&= \text{Cov}(N(t+1), N(t+s+1)) - \text{Cov}(N(t+1), N(t+s))
\end{aligned}
$$

For any $a < b$, we have

$$
\begin{aligned}
\text{Cov}(N(a), N(b)) &= \text{Cov}(N(a), N(a) + (N(b) - N(a))) \\
&= Var(N(a)) + \text{Cov}(N(a), N(b) - N(a)) \\
&= Var(N(a)) + 0 \\
&= \lambda a
\end{aligned}
$$

where $\text{Cov}(N(a), N(b) - N(a)) = 0$ because they are independent random variables.

Using this, we have

$$
\begin{aligned}
&\text{Cov}(N(t+1), N(t+s+1)) - \text{Cov}(N(t+1), N(t+s)) \\
&= \lambda(t+1) - \lambda(t+s) \\
&= \lambda(1-s)
\end{aligned}
$$

2. Let $\{N(t, t \geq 0)\}$ denote a Poisson process with rate λ and define $Y(t)$ to be the time from t until the next Poisson event.

   (a) Argue that $\{Y(t), t \geq 0\}$ is a stationary process.
   Solution: Starting at any time, a Poisson process will remain a Poisson process with the same parameters.

   (b) Compute $\text{Cov}(Y(t), Y(t+s))$.
   Solution: We have

$$
\begin{aligned}
\text{Cov}(Y(t), Y(t+s)) &= E(Y(t)(t+s)) \\
&= \int_0^\infty E(Y(t)Y(t+s)|Y(t) = y) f_{Y(t)}(y) dy \\
&= \int_0^\infty E(Y(t)Y(t+s)|Y(t) = y) \lambda e^{-\lambda y} dy \\
&= \int_0^\infty yE(Y(t+s)|Y(t) = y) \lambda e^{-\lambda y} dy
\end{aligned}
$$

3. Let $\{X(t), -\infty < t < \infty\}$ be a weakly stationary process having covariance function

$$R_X(s) = \text{Cov}(X(t), X(t+s))$$

(a) Show that

$$Var(X(t+s) - X(t)) = 2R_X(0) - 2R_X(t)$$

Solution: We have

$$\begin{aligned} Var(X(t+s) - X(t)) &= Var(X(t+s) - 2\text{Cov}(X(t), X(t+s)) + Var(X(t))) \\ &= R_X(0) - 2R_X(s) + R_X(0) \\ &= 2R_X(0) - 2R_X(s) \end{aligned}$$

(b) If $Y(t) = X(t+1) - X(t)$ show that $\{Y(t), -\infty < t < \infty\}$ is also weakly stationary having covariance function

$$R_Y(s) = 2R_X(s) - R_X(s-1) - R_X(s+1)$$

Directly calculate the covariance function:

$$\begin{aligned} \text{Cov}(Y(t), Y(t+s)) &= \text{Cov}(X(t+1) - X(t), X(t+s+1) - X(t+s)) \\ &= \text{Cov}(X(t+1), X(t+s+1)) - \text{Cov}(X(t+1), X(t+s)) \\ &\quad - \text{Cov}(X(t), X(t+s+1)) - \text{Cov}(X(t), X(t+s)) \\ &= R_X(s) - R_X(s-1) - R_X(s+1) + R_X(s) \\ &= 2R_X(s) - R_X(s-1) - R_X(s+1) \end{aligned}$$

4. Let Y_1 and Y_2 be independent standard normal random variables and for some constant w, set

$$X(t) = Y_1 \cos(wt) + Y_2 \sin(wt)$$

for $-\infty < t < \infty$. Show that $\{X(t)\}$ is a weakly stationary process.
Solution: We must check that $\text{Cov}(X(t), X(t+s))$ does not depend on t. Since the Y_i are standard normal random variables:

$$\begin{aligned} E(X(t)) &= E(Y_1 \cos(wt) + Y_2 \sin(wt)) \\ &= E(Y_1)\cos(wt) + E(Y_2)\sin(wt) \\ &= 0 \\ E(X(t+s)) &= 0 \end{aligned}$$

Hence:

$$\begin{aligned} \text{Cov}(X(t), X(t+s)) &= \text{Cov}(Y_1 \cos(wt) + Y_2 \sin(wt), Y_1 \cos(w(t+s)) + Y_2 \sin(w(t+s))) \\ &= E((Y_1 \cos(wt) + Y_2 \sin(wt))(Y_1 \cos(w(t+s)) + Y_2 \sin(w(t+s)))) \\ &= E({Y_1}^2 \cos(wt)\cos(w(t+s))) + E(Y_1 Y_2 \cos(wt)\sin(w(t+s))) \\ &\quad + E(Y_2 Y_1 \sin(wt)\cos(w(t+s))) + E({Y_2}^2 \sin(wt)\sin(w(t+s))) \\ &= \cos(wt)\cos(w(t+s))E({Y_1}^2) + \cos(wt)\sin(w(t+s))E(Y_1 Y_2) \\ &\quad + \sin(wt)\cos(w(t+s))E(Y_2 Y_1) + \sin(wt)\sin(w(t+s))E({Y_2}^2) \\ &= \cos(wt)\cos(w(t+s)) + \sin(wt)\sin(w(t+s)) \end{aligned}$$

We can then use the following trigonometric identities:

$$\cos(wt + ws) = \cos(wt)\cos(ws) - \sin(wt)\sin(ws)$$
$$\sin(wt + ws) = \sin(wt)\sin(ws) + \cos(wt)\cos(ws)$$

to get:

$$\begin{aligned}
&\cos(wt)\cos(w(t+s)) + \sin(wt)\sin(w(t+s)) \\
&\quad = \cos(wt)(\cos(wt)\cos(ws) - \sin(wt)\sin(ws)) + \sin(wt)(\sin(wt)\sin(ws) + \cos(wt)\cos(ws)) \\
&\quad = \cos(ws)
\end{aligned}$$

which indeed does not depend on t. Hence, it is a weakly stationary process.

5. Let $\{X(t), -\infty < t < \infty\}$ be weakly stationary with covariance function

$$R(s) = \mathrm{Cov}(X(t), X(t+s))$$

and let $\tilde{R}(w)$ denote the PSD of the process.

(a) Show that

$$\tilde{R}(w) = \tilde{R}(-w)$$

Note that it can be shown that $R(s) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \tilde{R}(w)e^{iws}dw$.
Solution: We will show that the autocorrelation function is even. Then it would follow that the PSD is symmetric about 0 by taking the Fourier transform of both sides.

We have that

$$R(s) = \int_{-\infty}^{\infty} X(t)X(t+s)dt$$

and substituting $t = \tau - s$ yields

$$\begin{aligned}
&\int_{-\infty}^{\infty} X(t)X(t+s)dt \\
&= \int_{-\infty}^{\infty} X(\tau - s)X(\tau)d\tau \\
&= \int_{-\infty}^{\infty} X(\tau)X(\tau - s)d\tau \\
&= R(-s)
\end{aligned}$$

so indeed, the autocorrelation function is symmetric.

(b) Use the preceding to show that

$$\int_{-\infty}^{\infty} \tilde{R}(w)dw = 2\pi E(X^2(t))$$

Solution: We had that

$$R(s) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \tilde{R}(w)e^{iws}dw$$

Plug in $s = 0$ to get:

$$\begin{aligned} R(0) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} \tilde{R}(w)e^{iw*0}dw \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty} \tilde{R}(w)dw \end{aligned}$$

Note that the left side of the equation is

$$\begin{aligned} R(0) &= \mathrm{Cov}(X(t), X(t+0)) \\ &= Var(X(t)) \\ &= E(X^2(t)) \end{aligned}$$

Hence:

$$\begin{aligned} E(X^2(t)) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} \tilde{R}(w)dw \\ 2\pi E(X^2(t)) &= \int_{-\infty}^{\infty} \tilde{R}(w)dw \end{aligned}$$

Conditional Exp, Martingale, Optional Stopping Thm, copied straight from yellow book...needs total revision

In this chapter, concepts that are a little beyond the scope of EE 126, and within the scope of EE 226, are explored:

Conditional Expectation

We first define a few things:

1. A random variable $S : \Omega \to \mathbb{R}$ is said to be *integrable* if:

$$\int_{\Omega} |S| dP < \infty$$

2. The *expectation* of a random variable X is given by

$$E(X) = \int_{\Omega} X dP$$

3. **Sigma Algebra**

   The *sigma algebra* of a (not necessarily finite) collection of subsets $A_1, \cdots, A_n$ of Ω (denoted $\sigma(A_1, \cdots, A_n)$) is defined to be the set containing Ω, the empty set $\emptyset$, each A_i, as well as their unions, intersections, and complements, so that it is closed under complement, closed under unions, and closed under intersections.

   Some examples are as follows: let $\Omega = [0, 1)$. Then

   (a) $\sigma(\left[0, \frac{1}{3}\right)) = \{\Omega, \emptyset, \left[0, \frac{1}{3}\right), \left[\frac{1}{3}, 1\right)\}$
   (b) $\sigma(\left[0, \frac{1}{3}\right), \left[\frac{1}{3}, \frac{2}{3}\right)) = \{\Omega, \emptyset, \left[0, \frac{1}{3}\right), \left[\frac{1}{3}, \frac{2}{3}\right), \left[\frac{1}{3}, 1\right), \left[\frac{2}{3}, 1\right), \left[\frac{2}{3}, 1\right), \left(0, \frac{2}{3}\right]\}$
   (c) $\sigma(\left[0, \frac{1}{4}\right), \left[\frac{1}{2}, \frac{3}{4}\right))$ (exercise)
   (d) $\sigma(\{\frac{1}{2}\})$ (exercise)

We shall consider different cases depending on what it is we condition on:

1. **Conditioning on an Event** This is the simplest case, and one that we may be most familiar with.

   For any integrable random variable Y and any event $B \in \mathcal{F}$ such that $P(B) \neq 0$, the *conditional expectation* of S given B is defined by

$$E(S|B) = \frac{1}{P(B)} \int_B S dP$$

   For example, suppose three fair coins are tossed. If the first coin lands heads, we are rewarded 10 dollars. If the second coin lands heads, we get 20 dollars, and the third coin, 30 dollars. Let the total sum of our dollars be denoted by S. We desire to find the expected value of S given that two of the coins have landed heads up, i.e $E(S|B)$ where B is the following event:

$$B = \{HHT, HTH, THH\}$$

The corresponding sums are then:

$$S(HHT) = 10 + 20 = 30$$
$$S(HTH) = 10 + 30 = 40$$
$$S(THH) = 20 + 30 = 50$$

and each sum occurs with probability $\frac{1}{2^3} = \frac{1}{8}$.

Plugging into the formula, we have:

$$\begin{aligned} E(S|B) &= \frac{1}{P(B)} \int_B S dP \\ &= \frac{1}{\frac{3}{8}}(30 + 40 + 50) * \frac{1}{8} \\ &= 40 \end{aligned}$$

(a) Show that $E(S|\Omega) = E(S)$.
Solution: Using the formula as above,

$$\begin{aligned} E(S|\Omega) &= \frac{1}{P(\Omega)} \int_\Omega S dP \\ &= \int_\Omega S dP \\ &= E(S) \end{aligned}$$

(b) Define

$$\mathbb{1}_A(\omega) = \begin{cases} 1 \text{ if } \omega \in A \\ 0 \text{ else} \end{cases}$$

Show that $E(\mathbb{1}_A|B) = P(A|B)$ where $P(A|B) = \frac{P(A\cap B)}{P(B)}$, the *conditional probability* of A given B.

Solution:

$$\begin{aligned} E(\mathbb{1}_A|B) &= \frac{1}{P(B)} \int_B \mathbb{1}_A dP \\ &= \frac{1}{P(B)} \int_{A\cap B} \mathbb{1}_A dP \\ &= \frac{P(A \cap B}{P(B)} \\ &= P(A|B) \end{aligned}$$

2. **Conditioning on a Discrete Random Variable** Suppose we have a discrete random variable X that takes on some possible values $x_1, x_2, x_3, \cdots$ with nonzero probability for each.

Then we define the conditional expectation of S given X to vary according to which value that X takes:

$$E(S|X)(\omega) = E(S|\{X = x_i\}) \text{ if } X(\omega) = x_i$$

for any $i = 1, 2, 3, \cdots$.

For example, we consider again the three fair coins with dollar reward values of $10, 20, 30$ respectively. What is the conditional expectation $E(S|X)$ of the total amount of money S obtained by the three coins, given the total number amount X obtained from the 10 and 20 coins only?

We first take note of the possible values that X can take:

$$\begin{aligned} X(TT) &= 0 + 0 = 0 \\ X(HT) &= 10 + 0 = 10 \\ X(TH) &= 0 + 20 = 20 \\ X(HH) &= 10 + 20 = 30 \end{aligned}$$

Now we condition on each of the above to calculate our desired quantity. For the first case, we have

$$\begin{aligned} E(S|X = 0) &= P(\text{last coin in Heads}) * S(TTH) + P(\text{last coin in Tails}) * S(TTT) \\ &= \frac{1}{2}(30) + \frac{1}{2}(0) \\ &= 15 \end{aligned}$$

Similarly for the other cases:

$$\begin{aligned} E(S|X = 10) &= P(\text{last coin in Heads}) * S(HTH) + P(\text{last coin in Tails}) * S(HTT) \\ &= \frac{1}{2}(40) + \frac{1}{2}(10) \\ &= 25 \end{aligned}$$

$$\begin{aligned} E(S|X = 20) &= P(\text{last coin in Heads}) * S(THH) + P(\text{last coin in Tails}) * S(THT) \\ &= \frac{1}{2}(50) + \frac{1}{2}(20) \\ &= 35 \end{aligned}$$

$$\begin{aligned} E(S|X = 30) &= P(\text{last coin in Heads}) * S(HHH) + P(\text{last coin in Tails}) * S(HHT) \\ &= \frac{1}{2}(60) + \frac{1}{2}(30) \\ &= 45 \end{aligned}$$

Hence we have overall:

$$E(S|X)(\omega) = \begin{cases} 15 \text{ if } X(\omega) = 0 \\ 25 \text{ if } X(\omega) = 10 \\ 35 \text{ if } X(\omega) = 20 \\ 45 \text{ if } X(\omega) = 30 \end{cases}$$

Now we will look at another example: we have $\Omega = (0, 1]$ with the σ-field of Borel sets, $S(x) = 2x^2$, P the Lebesgue measure on $[0, 1]$ and the following function of:

$$\eta(x) = \begin{cases} 1 \text{ if } x \in [0, \frac{1}{3}] \\ 2 \text{ if } x \in (\frac{1}{3}, \frac{2}{3}] \\ 0 \text{ if } x \in (\frac{2}{3}, 1] \end{cases}$$

We want to find $E(S|\eta)$.

For the case where $x \in [0, \frac{1}{3}]$, we have:

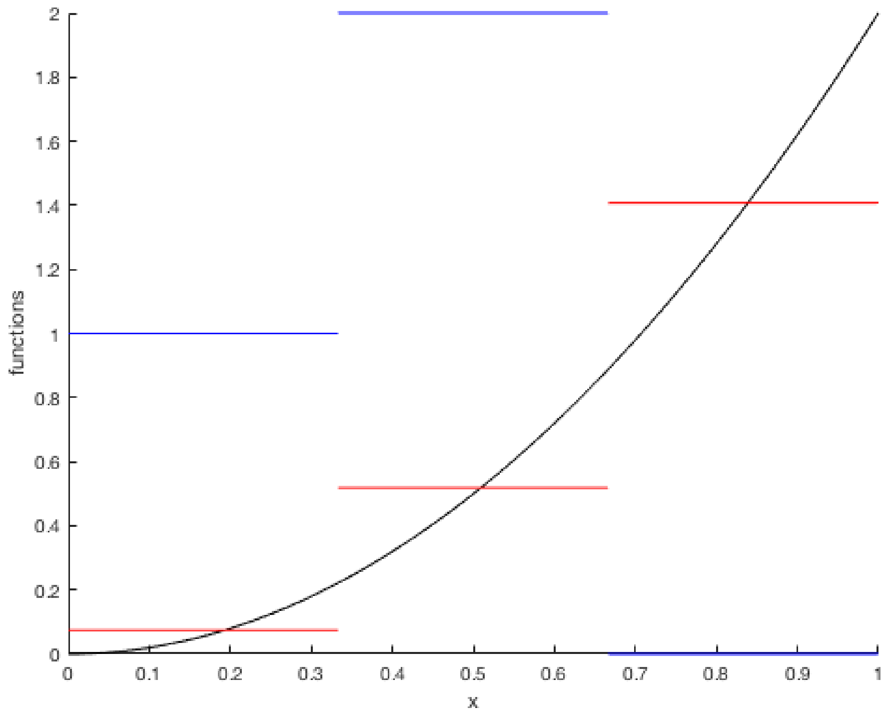
$$\begin{aligned} E(S|\eta)(x) &= E(S|[0, \frac{1}{3}]) \\ &= \frac{1}{\frac{1}{3}} \int_0^{\frac{1}{3}} S(x)dx \\ &= 3 \int_0^{\frac{1}{3}} 2x^2 dx \\ &= \frac{2}{27} \end{aligned}$$

Similarly for the other cases,

$$\begin{aligned} E(S|\eta)(x) &= E(S|(\frac{1}{3}, \frac{2}{3}]) \\ &= \frac{1}{\frac{1}{3}} \int_{\frac{1}{3}}^{\frac{2}{3}} S(x)dx \\ &= 3 \int_{\frac{1}{3}}^{\frac{2}{3}} 2x^2 dx \\ &= \frac{14}{27} \end{aligned}$$

$$\begin{aligned} E(S|\eta)(x) &= E(S|(\frac{2}{3}, 1]) \\ &= \frac{1}{\frac{1}{3}} \int_{\frac{2}{3}}^{1} S(x)dx \\ &= 3 \int_{\frac{2}{3}}^{1} 2x^2 dx \\ &= \frac{38}{27} \end{aligned}$$

We have the following visualization, where $S(x)$ is in black, $\eta(x)$ is in blue, and $E(S|\eta)$ is in red.



(a) Show that if η is a constant function, then $E(S|\eta)$ is constant and equal to $E(S)$.
Solution: Let $\eta(x) = c$, where c is a scalar belonging to $\mathbb{R}$. Then, only one event can occur: $\{\eta = c\}$, which means that it is the entire Ω set.

We then have

$$\begin{aligned} E(S|\eta)(\omega) &= E(S|\{\eta = c\}) \\ &= E(S|\Omega) \\ &= E(S) \end{aligned}$$

(b) Show that we have

$$E(\mathbb{1}_A|\mathbb{1}_B)(\omega) = \begin{cases} P(A|B) \text{ if } \omega \in B \\ P(A|\overline{B}) \text{ else} \end{cases}$$

when $0 < P(B) < 1$.
Solution: By definition,

$$\mathbb{1}_B = \begin{cases} 1 \text{ if } \omega \in B \\ 0 \text{ else} \end{cases}$$

Then by a previous exercise,

$$\begin{aligned} &E(\mathbb{1}_A|\mathbb{1}_B)(\omega) \\ &= E(\mathbb{1}_A|B) \\ &= P(A|B) \end{aligned}$$

if $\omega \in B$ and

$$\begin{aligned}&E(\mathbb{1}_A|\mathbb{1}_B)(\omega)\\&= E(\mathbb{1}_A|B)\\&= P(A|B)\end{aligned}$$

if $\omega \notin B$.

(c) Suppose that X is a discrete random variable. Show that

$$E(E(S|X)) = E(S)$$

Solution: Note that for any event B,

$$\begin{aligned}&\int_B E(S|X)dP\\&= \int_B (\frac{1}{P(B)} \int_B SdP)dP\\&= \int_B SdP\end{aligned}$$

Just as how we first defined conditioning on a discrete random variable, we know that X can take several values x_n for $n = 1, 2, \cdots$

$$\begin{aligned}E(E(S|X)) &= \int_\Omega E(S|X)dP\\&= \sum_n \int_{\{X=x_n\}} E(S|\{X = x_n\})dP\\&= \sum_n \int_{\{X=x_n\}} SdP\\&= \int_\Omega SdP\\&= E(S)\end{aligned}$$

3. **Conditioning on an Arbitrary Random Variable** We will use the following definition when conditioning on arbitrary random variables:

Let S be an integrable random variable and let X be an arbitrary random variable. Then the *conditional expectation* of S given X is defined to be the random variable $E(S|X)$ such that the two conditions below hold:

(a) $E(S|X)$ is $\sigma(X)$- measurable.

(b) For any $A \in \sigma(X)$,

$$\int_A E(S|X)dP = \int_A SdP$$

As a bit of an insight into how these cases arrive, we shall prove the following claim, which essentially looks into the case where the arbitrary random variable is a discrete random variable:

Claim: If S is an integrable random variable and X is a discrete random variable, then the above two conditions hold.
Proof: Assume X takes on the values $x_1, x_2, \cdots$. Then the σ-field is generated by the events $\{X = x_n\}, n = 1, 2, \cdots$.

Thus, every $A \in \sigma(n)$ is a countable union of sets of the form $\{X = x_n\}$, each of which are pairwise disjoint. It then follows that:

$$\int_A E(S|X)dP = \int_A SdP$$

Now since $E(S|X)$ is constant on each of the sets of such a form, it must be $\sigma(n)$-measurable. Then we have

$$\begin{aligned}&\int_{\{X=x_n\}} E(S|X)dP \\ &= \int_{\{X=x_n\}} E(S|\{X = x_n\})dP \\ &= \int_{\{X=x_n\}} SdP\end{aligned}$$

for each $n = 1, 2, \cdots$. This concludes our proof.

We shall also define the *conditional probability* of an event $A \in \mathcal{F}$ given X by the following:

$$P(A|X) = E(\mathbb{1}_A|X)$$

Now consider the following example: suppose we had $\Omega = (0, 1]$ with the σ-field of Borel sets, P the Lebesgue measure on $(0, 1]$, $S(x) = 5x^3$, and the function

$$\eta(x) = \begin{cases} 3 \text{ if } x \in (0, \frac{1}{3}] \\ x \text{ if } x \in (\frac{1}{3}, 1] \end{cases}$$

We desire to find $E(S|\eta)$. In order to do so, we must ensure that the conditions of the definition are satisfied.

First we must describe $\sigma(\eta)$. For $x \in (\frac{1}{3}, 1]$, any Borel set is measurable because the mapping function is nonconstant. On the other hand, when $x \in (0, \frac{1}{3}]$, the mapping function is constant, so not every Borel set can be measured.

Hence we describe the sigma field: for every $B \subset (\frac{1}{3}, 1], (0, \frac{1}{3}] \cup B \in \sigma(\eta)$. Now the first condition of the definition is satisfied.

Now in order for $E(S|\eta)$ to be measurable, it must be constant on $(0, \frac{1}{3}]$. So

$$\begin{aligned}
E(S|\eta)(x) &= E(S|(0, \frac{1}{3}]) \\
&= \frac{1}{P(0, \frac{1}{3}]} \int_0^{\frac{1}{3}} S(x)dx \\
&= 3 * \int_0^{\frac{1}{3}} 5x^3 dx \\
&= 3 * \frac{5}{4} \left(\frac{1}{3}\right)^4 \\
&= \frac{5}{108}
\end{aligned}$$

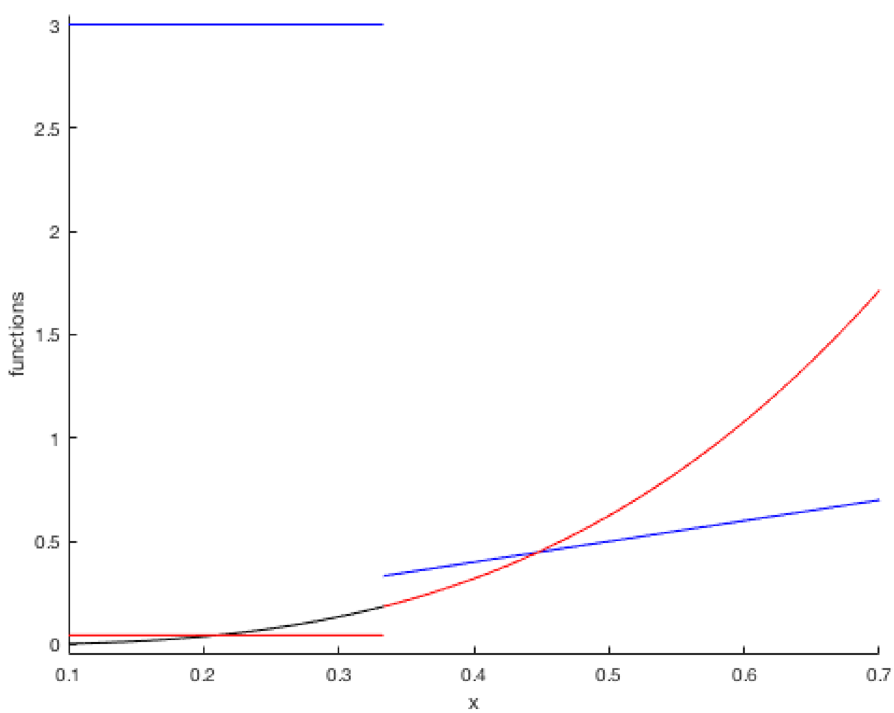
Now we satisfy the second condition in the interval $(0, \frac{1}{3}]$.

Furthermore, if $E(S|\eta) = S$ on the interval $(\frac{2}{3}, 1]$, then we certainly satisfy the second condition in the interval $(\frac{2}{3}, 1]$, and we can integrate over any Borel set in the interval $(\frac{2}{3}, 1]$.

Hence our final answer is

$$E(S|\eta) = \begin{cases} \frac{5}{108} \text{ if } x \in (0, \frac{1}{3}] \\ 5x^3 \text{ if } x \in (\frac{2}{3}, 1] \end{cases}$$

Overall, we get the graph below. $S(x)$ is in black, $\eta(x)$ is in blue, and $E(S|\eta)$ is in red.



(a) Let $\Omega = [0,1]$ with the Lebesgue measure as in above. We have $S(x) = 4x^2$ and $\eta(x) = 1 - |2x-1|$. Find the conditional expectation $E(S|\eta)$.
Solution: We observe that $\eta(x)$ is symmetric around $x = \frac{1}{2}$:

$$\eta(x) = \eta(1-x)$$

for any $x \in [0,1]$.

Then the only sets that are measurable are the Borel sets $A \subset [0,1]$ that are symmetric about $\frac{1}{2}$:

$$A = 1 - A$$

where

$$1 - A := \{1 - x : x \in A\}$$

So $\sigma(\eta)$ consists of all Borel sets that are symmetric about $x = \frac{1}{2}$. With this, we satisfy the first condition of the definition.

Now we must calculate $E(S|\eta)$. In order to be $\sigma(\eta)$-measurable, is must be symmetric about $x = \frac{1}{2}$:

$$E(S|\eta)(x) = E(S|\eta)(1-x)$$

Now compute the integral, so that we can find an expression for $E(S|\eta)$ such that the second condition holds:

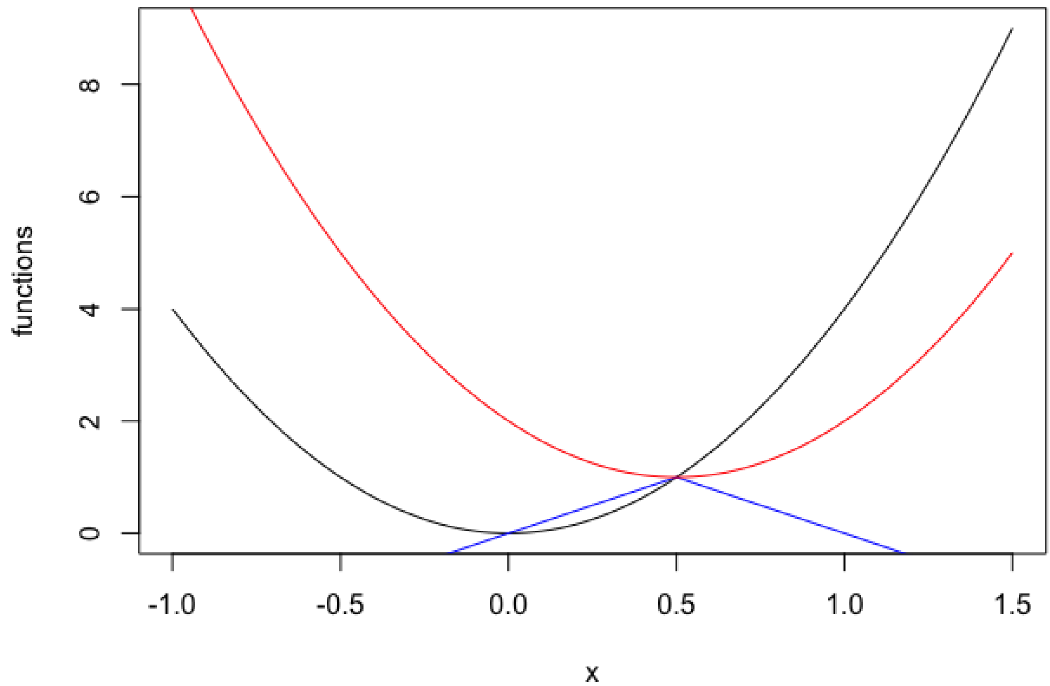
$$\begin{aligned}
\int_A E(S|\eta)(x)dx &= \int_A S(x)dx \\
&= \int_A 4x^2 dx \\
&= \int_A 2x^2 dx + \int_A 2x^2 dx \\
&= \int_A 2x^2 dx + \int_{1-A} 2(1-x)^2 dx \\
&= \int_A 2x^2 dx + \int_A 2(1-x)^2 dx \\
&= \int_A 2(x^2 + (1-x)^2) dx
\end{aligned}$$

because of the symmetry about $\frac{1}{2}$.

Matching up terms, we see that

$$E(S|\eta)(x) = 2(x^2 + (1-x)^2)$$

We have the following visualizations of our results, where $S(x)$ is in black, $\eta(x)$ is in blue, and $E(S|\eta)$ is in red. Note especially the symmetry of $E(S|\eta)$ around $\frac{1}{2}$ and the similar form it takes as $S(x)$.



(b) Let Ω be the unit square $[0,1]\times[0,1]$, with the σ-field of Borel sets and P the Lebesgue measure on $[0,1]\times[0,1]$. Suppose that S and η are random variables on Ω with the joint density

$$f_{S,\eta}(x,y) = \begin{cases} x+y \text{ for } x,y \in [0,1]\times[0,1] \\ 0 \text{ else} \end{cases}$$

Show that

$$E(S|\eta) = \tfrac{2+3\eta}{3+6\eta}$$

Solution: Note that any Borel set B in $[0,1]$ becomes two dimensional:

$$\{\eta \in B\} = [0,1]\times B$$

So we have

$$\begin{aligned}\int_{[0,1]\times B} S(x)dP &= \int_B \int_{[0,1]} x f_{S,\eta}(x,y)dxdy \\ &= \int_B \int_{[0,1]} x(x+y)dxdy \\ &= \int_B \int_0^1 (x^2+xy)dxdy \\ &= \int_B (\frac{1}{3}+\frac{1}{2}y)dy\end{aligned}$$

We must verify that our given $E(S|\eta)$ satisfies condition 2 of the definition:

$$\begin{aligned}\int_{[0,1]\times B} E(S|\eta)dP &= \int_B \int_{[0,1]} \frac{2+3y}{3+6y} f_{S,\eta}(x,y)dxdy \\ &= \int_B \frac{2+3y}{3+6y} \int_0^1 (x+y)dxdy \\ &= \int_B \frac{2+3y}{3+6y}(\frac{1}{2}+y)dy \\ &= \int_B (\frac{1}{3}+\frac{1}{2}y)dy\end{aligned}$$

which equals $\int_{[0,1]\times B} S(x)dP$ as desired.

(c) Again let Ω be the unit square $[0,1]\times[0,1]$, with the σ-field of Borel sets and P the Lebesgue measure on $[0,1]\times[0,1]$. Suppose that S and η are random variables on Ω with the joint density

$$f_{S,\eta}(x,y) = \begin{cases} x^2+y^2 \text{ for } x,y \in [0,1]\times[0,1] \\ 0 \text{ else}\end{cases}$$

Find $E(S|\eta)$.

Solution: Assuming the same $\sigma(\eta)$-measurable form (hence satisfactory of the first condition of the definition) and using the same approach as in the previous exercise, we will first calculate the following integral:

$$\begin{aligned}\int_{[0,1]\times B} SdP &= \int_B \int_{[0,1]} x f_{S,\eta}(x,y)dxdy \\ &= \int_B \int_0^1 x(x^2+y^2)dxdy \\ &= \int_B \int_0^1 x^3+xy^2)dxdy \\ &= \int_B (\frac{1}{4}+\frac{1}{2}y^2)dy\end{aligned}$$

And we want an expression for $E(S|\eta)$ such that the second condition of the definition is satisfied:

$$\begin{aligned}\int_{[0,1]\times B} E(S|\eta)dP &= \int_B \int_{[0,1]} E(S|\eta)(y) f_{S,\eta}(x,y)dxdy \\ &= \int_B E(S|\eta)(y) \int_0^1 (x^2+y^2)dxdy \\ &= \int_B E(S|\eta)(y)(\frac{1}{3}+y^2)dy\end{aligned}$$

Matching up the integrands, we get:

$$\begin{aligned}E(S|\eta)(y)(\tfrac{1}{3}+y^2) &= \tfrac{1}{4}+\tfrac{1}{2}y^2 \\ E(S|\eta)(y) &= \frac{\frac{1}{4}+\frac{1}{2}y^2}{\frac{1}{3}+y^2} \\ &= \frac{3+6y^2}{4+12y^2}\end{aligned}$$

Hence we get that

$$E(S|\eta) = \tfrac{3+6\eta^2}{4+12\eta^2}$$

(d) Let Ω be the unit disk $\{(x,y)|x^2+y^2 \leq 1\}$ with the σ-field of Borel sets and P the Lebesgue measure on the disk normalized so that $P(\Omega)=1$. This means that for any Borel set $A \subset \Omega$,

$$P(A) = \frac{1}{\pi}\int\int_A dxdy$$

Suppose that S and η are projections onto the x and y axes, respectively, for any $(x,y) \in \Omega$. That is:

$$S(x,y) = x$$
$$\eta(x,y) = y$$

Find $E(S^2|\eta)$.
Solution: First observe that the random variables S and η have uniform joint distribution over the unit disk $\Omega = \{(x,y)|x^2+y^2 \leq 1\}$ with density

$$f_{S,\eta}(x,y) = \begin{cases} \frac{1}{\pi} \text{ if } x^2+y^2 \leq 1 \\ 0 \text{ else} \end{cases}$$

Once again, we compute integrals as follows:

$$\begin{aligned}
\int_{\{\eta\in B\}} S^2 dP &= \int_B\int_{\text{circle}} x^2 f_{S,\eta}(x,y)dxdy \\
&= \frac{1}{\pi}\int_B\int_{-\sqrt{1-y^2}}^{\sqrt{1+y^2}} x^2 dxdy \\
&= \frac{1}{\pi}\int_B \frac{1}{3}((1-y^2)^{\frac{3}{2}} + (1-y^2)^{\frac{3}{2}}) \\
&= \frac{2}{3\pi}\int_B (1-y^2)^{\frac{3}{2}} dy
\end{aligned}$$

And on the other hand, we get

$$\begin{aligned}
\int_{\{\eta\in B\}} E(S^2|\eta)dP &= \int_B\int_{\text{circle}} E(S^2|\eta)(y) f_{S,\eta}(x,y)dxdy \\
&= \frac{1}{\pi}\int_B E(S^2|\eta)(y)\int_{-\sqrt{1-y^2}}^{\sqrt{1+y^2}} dxdy \\
&= \frac{2}{\pi}\int_B E(S^2|\eta)(y)(1-y^2)^{\frac{1}{2}} dy
\end{aligned}$$

We then have that

$$\begin{aligned}
E(S^2|\eta)(y) &= \tfrac{1}{3}(1-y^2) \\
\implies E(S^2|\eta) &= \tfrac{1}{3}(1-\eta^2)
\end{aligned}$$

4. **Conditioning on a σ-field** Let S be an integrable random variable on a probability space $(\Omega, \mathcal{F}, P)$. Further let $\mathcal{G}$ be a σ-field contained in F. Then the *conditional expectation* of S given $\mathcal{G}$ is $E(S|\mathcal{G})$, defined to be the random variable such that the following two conditions hold:

   (a) $E(S|\mathcal{G})$ is $\mathcal{G}$-measurable
   (b) For any $A \in \mathcal{G}$,

   $$\int_A E(S|\mathcal{G})dP = \int_A SdP$$

   Radon-Nikodym Theorem: Let $(\Omega, \mathcal{F}, P)$ be a probability space and $\mathcal{G}$ be a σ-field contained in $\mathcal{F}$. Then for any random variable S, there exists a $\mathcal{G}$- measurable random variable U such that

   $$\int_A SdP = \int_A UdP$$

   (a) Show that if $\mathcal{G} = \{\emptyset, \Omega\}$ then $E(S|\mathcal{G}) = E(S)$ almost surely.
   Solution: Since $\mathcal{G} = \{\emptyset, \Omega\}$, any constant random variable is $\mathcal{G}$-measurable.

   We also have that

   $$\int_\Omega \eta dP = E(\eta) = \int_\Omega E(\eta)dP$$

   and

   $$\int_\emptyset \eta dP = 0 = \int_\emptyset E(\eta)dP$$

   Hence it follows that $E(\eta|\mathcal{G}) = E(\eta)$.
   (b) Show that if $B \in \mathcal{G}$, then

   $$E(E(S|\mathcal{G})|B) = E(S|B)$$

   Solution: By the definition and the fact that $B \in \mathcal{G}$. we have that

   $$\int_B E(S|\mathcal{G})dP = \int_B SdP$$

   It then follows that

   $$\begin{aligned} E(E(S|\mathcal{G})|B) &= \frac{1}{P(B)} \int_B E(S|\mathcal{G})dP \\ &= \frac{1}{P(B)} \int_B SdP \\ &= E(S|B) \end{aligned}$$

   and so we reach our desired conclusion.

The following is a list of some useful properties of the conditional expectation. a, b are real scalars, S, U are integrable random variables on a probability space $(\Omega, \mathcal{F}, P)$, and $\mathcal{G}, \mathcal{H}$ are σ-fields on Ω contained in $\mathcal{F}$.

(a) $E(aS + bU|\mathcal{G}) = aE(S|\mathcal{G}) + bE(U|\mathcal{G})$

(b) $E(E(S|\mathcal{G})) = E(S)$

(c) $E(SU|\mathcal{G}) = SE(U|\mathcal{G})$ if S is $\mathcal{G}$-measurable.

(d) $E(S|\mathcal{G}) = E(S)$ if S is independent of $\mathcal{G}$

(e) $E(E(S|\mathcal{G})|\mathcal{H}) = E(S|\mathcal{H})$ if $\mathcal{H} \subset \mathcal{G}$

(f) If $S \geq 0$ then $E(S|\mathcal{G}) \geq 0$

Jensen's Inequality: Let $\psi : \mathbb{R} \to \mathbb{R}$ be a convex function and let S be an integrable random variable on the probability space $(\Omega, \mathcal{F}, P)$ such that $\psi(S)$ is also integrable. Then

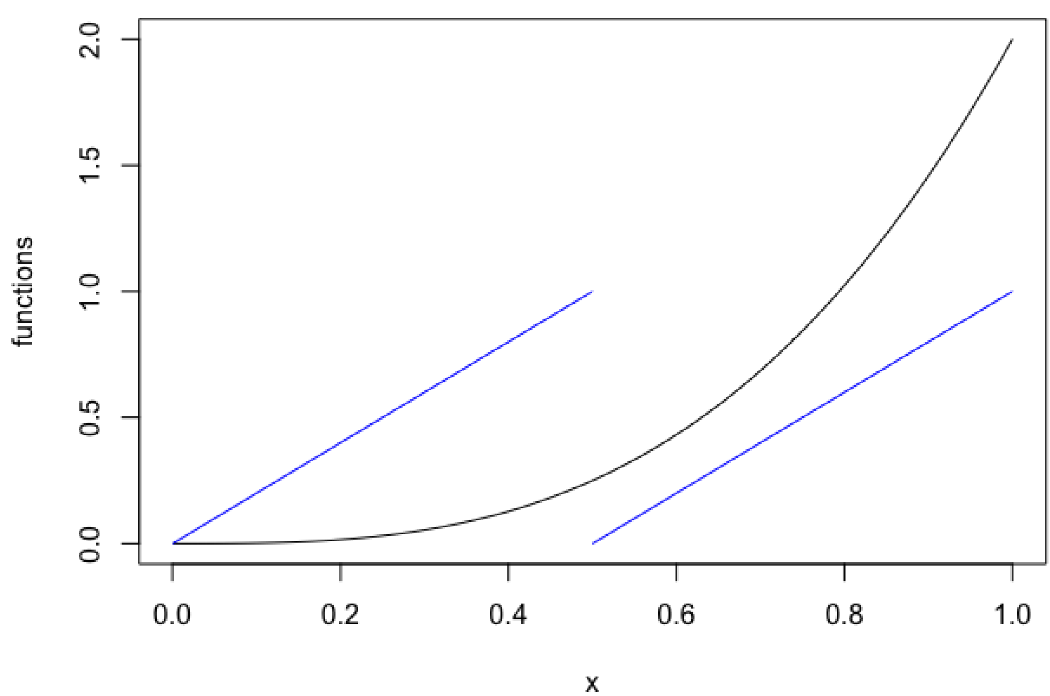
$$\psi(E(S|\mathcal{G})) \leq E(\psi(S)|\mathcal{G}) \text{ almost surely}$$

for all σ-fields $\mathcal{G} \subset \mathcal{F}$ on Ω.

(a) Let $\Omega = [0, 1)$ with the σ-field of Borel sets and the Lebesgue measure on $[0, 1)$. Take this to be the probability space. Also define $S(x) = 2x^2$ and

$$\eta(x) = \begin{cases} 2x \text{ for } 0 \leq x < \frac{1}{2} \\ 2x - 1 \text{ for } \frac{1}{2} \leq x < 1 \end{cases}$$

Find the conditional expectation $E(S|\eta)$.
Solution: Observe the following graph, where $S(x)$ is in black and $\eta(x)$ is in blue:



We must first find which Borel sets are $\sigma(\eta)$-measurable. Note that the sigma field $\sigma(\eta)$ must consist of Borel sets of the form $B \cup (B + \frac{1}{2})$ for some Borel sets $B \subset (0, \frac{1}{2}]$, according to the graph above.

So in order to satisfy the definition, we must seek a random variable U such that for each Borel set $B \subset (0, \frac{1}{2}]$,

$$\int_{B\cup(B+\frac{1}{2})} S(x)dx = \int_{B\cup(B+\frac{1}{2})} U(x)dx$$

Then we can say $U = E(S|\eta)$.

The left side of the equality gives us

$$\begin{aligned}
\int_{B\cup(B+\frac{1}{2})} S(x)dx &= \int_{B\cup(B+\frac{1}{2})} 2x^2 dx \\
&= \int_B 2x^2 dx + \int_{(B+\frac{1}{2})} 2x^2 dx \\
&= \int_B 2x^2 dx + \int_B 2(x+\frac{1}{2})^2 dx \\
&= 2\int_B x^2 dx + 2\int_B (x+\frac{1}{2})^2 dx \\
&= 2\int_B (x^2 + (x+\frac{1}{2})^2) dx
\end{aligned}$$

As we said before, in order for U to be $\sigma(\eta)$-measurable, it must satisfy

$$U(x) = U(x + \tfrac{1}{2})$$

for every $x \in (0, \frac{1}{2}]$.

Then the right side of the equation becomes:

$$\begin{aligned}
\int_{B\cup(B+\frac{1}{2})} U(x)dx &= \int_B U(x)dx + \int_{(B+\frac{1}{2})} U(x)dx \\
&= \int_B U(x)dx + \int_B U(x+\frac{1}{2})dx \\
&= \int_B U(x)dx + \int_B U(x)dx \\
&= \int_B 2U(x)dx
\end{aligned}$$

by the property of $U(x)$ we described above.

Matching termwise between each side of the equality, we get

$$U(x) = x^2 + (x + \tfrac{1}{2})^2$$

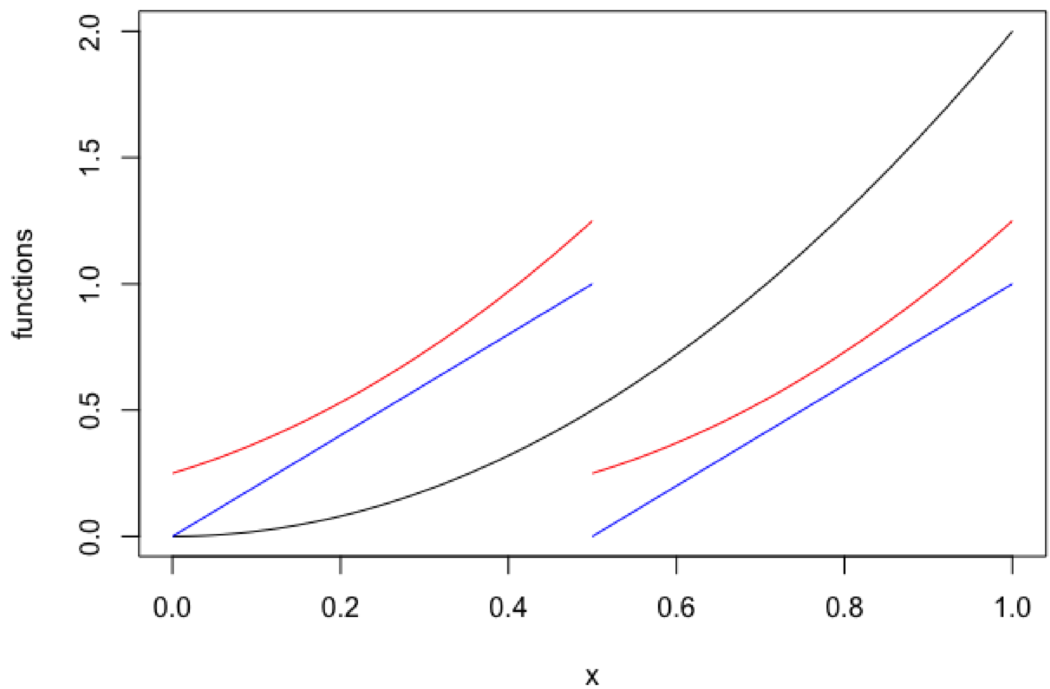
for every $x \in (0, \frac{1}{2}]$.

The case when $x \in (\frac{1}{2}, 1]$ can be analyzed similarly.

Overall, we obtain the following expression:

$$E(S|\eta)(x) = \begin{cases} x^2 + (x + \frac{1}{2})^2 \text{ for } x \in (0, \frac{1}{2}] \\ (x - \frac{1}{2})^2 + x^2 \text{ for } x \in (\frac{1}{2}, 1] \end{cases}$$

Finally, we end up with the following graph, now with $E(S|\eta)$ in red.



(b) Take $\Omega = [0, 1]$ with the σ-field of Borel sets and P the Lebesgue measure on $[0, 1]$. Let $\eta(x) = x(1 - x)$ for $x \in [0, 1]$. Show that

$$E(S|\eta)(x) = \frac{S(x)+S(1-x)}{2}$$

Solution: In order for the expectation to be $\sigma(\eta)$-measurable, all Borel sets $B \subset [0, 1]$ must have the following property:

$$B = 1 - B$$

where

$$1 - B := \{1 - x : x \in B\}$$

because we are given $\eta(x) = \eta(1 - x)$.

So for any such B that satisfies the above,

$$\begin{aligned}
\int_B S(x)dx &= \frac{1}{2}\int_B S(x)dx + \frac{1}{2}\int_B S(x)dx \\
&= \frac{1}{2}\int_B S(x)dx + \frac{1}{2}\int_{1-B} S(1-x)dx \\
&= \frac{1}{2}\int_B S(x)dx + \frac{1}{2}\int_B S(1-x)dx \\
&= \frac{1}{2}\int_B (S(x)+S(1-x))dx
\end{aligned}$$

So indeed we have $E(S|\eta)(x) = \frac{S(x)+S(1-x)}{2}$.

Martingales

We define a *filtration* to be a sequence of σ-fields $\mathcal{F}_1, \mathcal{F}_2, \mathcal{F}_3, \cdots$ on Ω such that

$$\mathcal{F}_1 \subset \mathcal{F}_2 \subset \mathcal{F}_3 \subset \cdots \subset \mathcal{F}$$

In the Martingale scenario, $\mathcal{F}_n$ represents our knowledge at time n.

For example, in a sequence of coin tosses $X_1, X_2, \cdots$, we take

$$F_n = \sigma(X_1, X_2, \cdots X_n)$$

Suppose A denotes the event that the first five tosses have at least two Heads. Then $A \in \mathcal{F}_5$ because it will be possible to determine whether A occurred or not after five tosses.

However, $A \notin \mathcal{F}_4$. With only four tosses made, we wouldn't be able to say with certainty whether two Heads have occurred in five tosses, especially if only one Head appeared among the four tosses that we know.

Now consider the following events:

1. $A = \{$the first Head is preceded by no more that ten tails$\}$
2. $B = \{$there is at least one Head in the sequence $X_1, X_2\}$
3. $C = \{$the first 100 tosses produce the same outcome$\}$
4. $D = \{$no more than two Heads and two Tails amont the first five tosses occur$\}$

For each of these events, find the smallest value of n such that the event is in F_n.

Solution: For A, the smallest such n is 11 since if there is a Head in any of the first 11 tosses, then we know that A has occurred. Otherwise, A cannot occur.

For B, the smallest n is undetermined since there is no way of knowing for sure when the first Heads will occur.

For C, $n = 100$. Similar to the example previously, any value of n smaller than 100 will give us insufficient information.

Note that event D cannot even happen, since a coin is bound to land either Heads or Tails. Hence, $n = 1$.

A sequence of random variables $X_1, X_2, \cdots$ is *adapted* to a filtration $\mathcal{F}_1, \mathcal{F}_2, \cdots$ if X_n is $\mathcal{F}_n$-measurable for each $n = 1, 2, \cdots$.

For example, if $\mathcal{F}_n = \sigma(X_1, X_2, \cdots, X_n)$, then $X_1, X_2, \cdots$ is adapted to $\mathcal{F}_1, \mathcal{F}_2, \cdots$.

We will show that this is the smallest filtration such that the sequence is adapted to it. Suppose $\mathcal{G}_1, \mathcal{G}_2, \cdots$ is another filtration such that $X_1, X_2, \cdots$ is adapted to it. We need to show that $F_n \subset G_n$, which implies that we need to show that $\sigma(X_1, X_2, \cdots) \subset \mathcal{G}$, which implies that we need to show that $X_1, X_2, \cdots$ are $\mathcal{G}_n$-measurable.

Note that since $X_1, X_2, \cdots$ is adapted to the filtration $\mathcal{G}_1, \mathcal{G}_2, \cdots$, we have that X_n is $\mathcal{G}_n$-measurable for each n. And since $\mathcal{G}_1 \subset \mathcal{G}_2 \subset \cdots \subset G_n$, $X_1, X_2, \cdots X_n$ is $\mathcal{G}_n$-measurable for each n.

Hence, $\sigma(X_1, \cdots X_n)$ are $\mathcal{G}_n$-measurable for each n.

A sequence $X_1, X_2, \cdots$ of random variables is a *martingale* with respect to the filtration $\mathcal{F}_1, \mathcal{F}_2, \cdots$ if

1. X_1 is integrable for each $n = 1, 2, \cdots$
2. $X_1, X_2, \cdots$ is adapted to $\mathcal{F}_1, \mathcal{F}_2, \cdots$
3. $E(X_{n+1}|\mathcal{F}_n) = X_n$ almost surely for each $n = 1, 2, \cdots$

For example, let $X_1, X_2, \cdots$ be sequence of independent and integrable random variables such that $E(X_n) = 0 \forall n = 1, 2, \cdots$
Put $Y_n = X_1 + X_2 + \cdots + X_n$ and $\mathcal{F}_n = \sigma(X_1, \cdots X_n)$. Then Y_n is adapted the filtration $\mathcal{F}_n$ since Y_n is $\sigma(X_1, \cdots, X_n)$-measurable.

Furthermore Y_n is integrable because

$$\begin{aligned} E(|Y_n|) &= E(|X_1 + \cdots + X_n|) \\ &\leq E(|X_1|) + \cdots + E(|X_n|) \text{ by Triangle Inequality} \\ &< \infty \text{ since } X_i \text{ are integrable} \end{aligned}$$

Also,

$$\begin{aligned} E(Y_{n+1}|\mathcal{F}_n) &= E(X_{n+1}|\mathcal{F}_n) + E(X_n|\mathcal{F}_n) \\ &= E(X_{n+1}) + X_n \\ &= X_n \end{aligned}$$

where the second line comes from the fact that X_{n+1} is independent of $\mathcal{F}_n$ and Y_n is $\mathcal{F}_n$-measurable.

Thus, Y_n is a martingale with respect to $\mathcal{F}_n$.

As another example, let Y be an integrable random variable and let $\mathcal{F}_1, \mathcal{F}_2, \cdots$ be a filtration. Put $Y_n = E(Y|\mathcal{F}_n)$.

Observe that

$$\begin{aligned} |Y_n| &= |E(Y|\mathcal{F}_n)| \\ &\le E(|Y||\mathcal{F}_n) \text{ by Triangle Inequality} \end{aligned}$$

This implies that

$$\begin{aligned} E(Y_{n+1}|\mathcal{F}_n) &= E(E(Y|\mathcal{F}_{n+1})|\mathcal{F}_n) \\ &= E(Y|\mathcal{F}_n) \\ &= Y_n \end{aligned}$$

where the second equality comes from the fact that $\mathcal{F}_n \subset \mathcal{F}_{n+1}$ and by a property of conditional expectation.

Clearly Y_n are integrable and are adapted to the filtration $\mathcal{F}_1, \mathcal{F}_2, \cdots$.

1. $\{Y_n\}$ is a martingale with respect to $\mathcal{F}_n \Longrightarrow E(Y_1) = E(Y_2) = \cdots$
   Proof: Since Y_n is a martingale with respect to $\mathcal{F}_n$, we have $E(Y_{n+1}|\mathcal{F}_n) = Y_n \forall n = 1, 2, \cdots$

   Then $E(E(Y_{n+1}|\mathcal{F}_n)) = E(Y_n)$. By law of total expectation, the left had side of the equation is $E(Y_{n+1})$. The conclusion follows from induction.

1. Y_n is a martingale with respect to $\mathcal{F}_n \Longrightarrow Y_n$ is a martingale with respect to $\mathcal{G}_n = \sigma(X_1, \cdots, X_n)$.
   Proof: Recall from previous example that $\mathcal{G}_n \subset \mathcal{F}_n$.

   Since Y_n is a martingale with respect to $\mathcal{F}_n$, we have

$$\begin{aligned} E(Y_{n+1}|\mathcal{F}_n) &= Y_n \forall n = 1, 2, \cdots \\ \Longrightarrow & E(E(Y_{n+1}|\mathcal{F}_n)|\mathcal{G}_n) \\ &= E(Y_n|\mathcal{G}_n) \end{aligned}$$

   On the left hand side of the equality, we have $E(Y_{n+1}|\mathcal{G}_n)$ by a property of the conditional expectation. On the right hand side, we have Y_n since Y_n is $\mathcal{G}_n$-measurable.

2. Let Y_n be a symmetric random walk.

$$Y_n = X_1 + \cdots + X_n$$

   where $X_1, X_2, \cdots$ is a sequence of iid random variables such that

$$P(X_i = 1) = P(X_i = -1) = \tfrac{1}{2}$$

Show that ${Y_n}^2 - n$ is a martingale with respect to the filtration $\mathcal{F}_n = \sigma(X_1, X_2, \cdots X_n)$.

Solution: We shall first check integrability of the random variable ${Y_n}^2 - n$. Note that

$$ {Y_n}^2 - n = (X_1 + \cdots + X_n)^2 - n $$

is a function of $X_1, \cdots X_n$. Thus, it is $\sigma(X_1, \cdots, X_n)$- measurable. Because $\mathcal{F}_n = \sigma(X_1, \cdots, X_n)$, ${Y_n}^2 - n$ must be adapted to $\mathcal{F}_n$.

By the Triangle Inequality,

$$ \begin{aligned} |Y_n| &= |X_1 + \cdots + X_n| \\ &\leq |X_1| + \cdots |X_n| \\ &= n(1) \\ &= n \end{aligned} $$

by the definition of the indicator random variables X_i.

So we then have

$$ \begin{aligned} E({Y_n}^2 - n) &\leq |E({Y_n}^2)| + |n| \\ &= E({Y_n}^2) + n \\ &\leq E(n^2) + n \\ &= n^2 + n < \infty \end{aligned} $$

Indeed, ${Y_n}^2 - n$ is integrable.

Now we shall check the third Martingale property.

$$ \begin{aligned} E({Y_{n+1}}^2|\mathcal{F}_n) &= E(X_{n+1}^2 + 2X_{n+1}Y_n + {Y_n}^2|\mathcal{F}_n) \\ &= E({X_{n+1}}^2|\mathcal{F}_n) + 2E(X_{n+1}S_n|\mathcal{F}_n) + E({X_n}^2|\mathcal{F}_n) \end{aligned} $$

Note that since Y_n and ${Y_n}^2$ are $\mathcal{F}_n$-measurable and Y_{n+1} is independent of $\mathcal{F}_n$

$$ \begin{aligned} E({X_{n+1}}^2|\mathcal{F}_n) &= E({X_{n+1}}^2) \\ E(X_{n+1}Y_n|\mathcal{F}_n) &= Y_n E(X_{n+1}) \\ E({X_n}^2|\mathcal{F}_n) &= {X_n}^2 \end{aligned} $$

so we have

$$ \begin{aligned} &E({X_{n+1}}^2|\mathcal{F}_n) + 2E(X_{n+1}S_n|\mathcal{F}_n) + E({X_n}^2|\mathcal{F}_n) \\ &= E({X_{n+1}}^2) + 2S_n E(X_{n+1}) + {X_n}^2 \\ &= 1 + 2 * 0 + {X_n}^2 \\ &= 1 + {X_n}^2 \end{aligned} $$

So we overall have

$$\begin{aligned}E({X_{n+1}}^2 - n - 1|\mathcal{F}_n) &= 1 + {X_n}^2 - n - 1\\ &= {X_n}^2 - n\end{aligned}$$

which verifies the third Martingale property.

3. Let Y_n be a symmetric random walk and $\mathcal{F}_n$ the filtration defined as above. Show that

$$Z_n = (-1)^n \cos(\pi Y_n)$$

is a martingale with respect to $\mathcal{F}_n$
Solution: Because Z_n is a function of Y_n, which is $\mathcal{F}_n$-measurable for each n, it is also $\mathcal{F}_n$-measurable for each n.

Furthermore, Z_n is bounded by 1 in magnitude. Hence, it is an integrable random variable.

Now we must check the third and final property of being a Martingale process. Since Y_n is a random walk, it can be written recursively as:

$$Y_n = Y_{n-1} + X_n$$

Hence,

$$\begin{aligned}E(Z_{n+1}|\mathcal{F}_n) &= E((-1)^{n+1}\cos(\pi Y_{n+1})|\mathcal{F}_n)\\ &= (-1)^{n+1}E(\cos(\pi Y_{n+1})|\mathcal{F}_n)\\ &= (-1)^{n+1}E(\cos(\pi(Y_n + X_{n+1}))|\mathcal{F}_n)\\ &= (-1)^{n+1}E(\cos(\pi Y_n)\cos(\pi X_{n+1}) - \sin(\pi Y_n)\sin(\pi X_{n+1}))|\mathcal{F}_n)\\ &= (-1)^{n+1}E(\cos(\pi Y_n)\cos(\pi X_{n+1})|\mathcal{F}_n) - (-1)^{n+1}E(\sin(\pi Y_n)\sin(\pi X_{n+1})|\mathcal{F}_n)\end{aligned}$$

Similar to the previous problem, we have that X_{n+1} is independent of $\mathcal{F}_n$ and Y_n is $\mathcal{F}_n$-measurable.

Hence

$$\begin{aligned}&(-1)^{n+1}E(\cos(\pi Y_n)\cos(\pi X_{n+1})|\mathcal{F}_n) - (-1)^{n+1}E(\sin(\pi Y_n)\sin(\pi X_{n+1})|\mathcal{F}_n)\\ &= \cos(\pi Y_n)(-1)^{n+1}E(\cos(\pi X_{n+1})|\mathcal{F}_n) - \sin(\pi Y_n)(-1)^{n+1}E(\sin(\pi X_{n+1})|\mathcal{F}_n)\\ &= (-1)^{n+1}\cos(\pi Y_n)E(\cos(\pi X_{n+1})) - (-1)^{n+1}\sin(\pi Y_n)E(\sin(\pi X_{n+1}))\end{aligned}$$

We have that

$$\begin{aligned}E(\cos(\pi X_{n+1}) &= E(\cos(\pi)) = E(\cos(-\pi)) = E(-1) = -1\\ E(\sin(\pi X_{n+1}) &= E(\sin(\pi)) = E(\sin(-\pi)) = E(0) = 0\end{aligned}$$

So

$$
\begin{aligned}
&(-1)^{n+1}\cos(\pi Y_n)E(\cos(\pi X_{n+1})) - (-1)^{n+1}\sin(\pi Y_n)E(\sin(\pi X_{n+1})) \\
&= (-1)^{n+1}\cos(\pi Y_n)(-1) \\
&= (-1)^{n+2}\cos(\pi Y_n) \\
&= (-1)^{n}\cos(\pi Y_n) \\
&= Z_n
\end{aligned}
$$

which indeed verifies the third Martingale property.

A sequence $Y_1, Y_2, \cdots$ is a *supermartingale* with respect to a filtration $\mathcal{F}_1, \mathcal{F}_2, \cdots$ if the following hold:

1. Y_n is integrable for each $n = 1, 2, \cdots$
2. $Y_1, Y_2, \cdots$ is adapted to $\mathcal{F}_1, \mathcal{F}_2, \cdots$
3. $E(Y_{n+1}|\mathcal{F}_n) \leq Y_n$ almost surely for each of $n = 1, 2, \cdots$

It is likewise a *submartingale* with respect to the filtration when

$$E(Y_{n+1}|\mathcal{F}_n) \geq Y_n \text{ for each of } n = 1, 2, \cdots$$

For example, let Y_n is a sequence of integrable random variables such that Y_n is a martingale with respect to the filtration $\mathcal{F}_n$. Show then that ${Y_n}^2$ is a submartingale with respect to the same filtration.

We can consider the convex function $\psi(x) = x^2$. Since $Y_n = E(Y_{n+1}|\mathcal{F}_n)$, we can apply Jensen's Inequality:

$$
\begin{aligned}
{Y_n}^2 &= E(Y_{n+1}|\mathcal{F}_n)^2 \\
&\leq E(Y_{n+1}^2|\mathcal{F}_n)
\end{aligned}
$$

Hence, ${Y_n}^2$ is certainly a submartingale with respect to the filtration.

We will now attempt to fit martingales into the context of games and more specifically, to the Gambler's Ruin Problem.

Let $X_1, X_2, \cdots$ be a sequence of integrable random variables where X_i are winnings/losses per unit stake (you bet one dollar per turn) on game n.

The total winnings after n games are then

$$
\begin{aligned}
Y_n &= X_1 + \cdots + X_n \\
Y_0 &= 0
\end{aligned}
$$

Take $\mathcal{F}_n = \sigma(X_1, X_2, \cdots X_n)$ to be the filtration and let $\mathcal{F}_0 = \{\emptyset, \Omega\}$. So if n games have been played so far, $\mathcal{F}_n$ denote the amount of knowledge accumulated up to time n.

We say the game is:

1. *fair* if $E(Y_n|F_{n-1}) = Y_{n-1}$ (we expect that the fortune at time step n is, on average, the same as the fortune at time step $n-1$.)

2. *favorable* to you if $E(Y_n|F_{n-1}) \geq Y_{n-1}$

3. *unfavorable* to you if $E(Y_n|F_{n-1}) \leq Y_{n-1}$

Suppose you vary the stake to be α_n at game n. We can assume α_n is $\mathcal{F}_{n-1}$-measurable because when the time comes to decide your stake α_n, you will have already known the outcomes of the previous $n-1$ games.

A *gambling strategy* $\alpha_1, \alpha_2, \cdots$ with respect to the filtration $\mathcal{F}_1, \mathcal{F}_2, \cdots$ is a sequence of random variables such that α_n is $\mathcal{F}_{n-1}$-measurable.

Total winnings:

$$\begin{aligned} Z_n &= \alpha_1 X_1 + \cdots + \alpha_n X_n \\ &= \alpha_1(Y_1 - Y_0) + \cdots + \alpha_n(Y_n - Y_{n-1}) \end{aligned}$$

and $Z_0 = 0$.

Suppose $\alpha_1, \alpha_2, \cdots$ is a gambling strategy.

1. If $\alpha_1, \alpha_2, \cdots$ is a bounded sequence and $Y_0, Y_1, Y_2, \cdots$ is a martingale then $Z_0, Z_1, Z_2, \cdots$ is also a martingale. This means that a fair game turns into a fair one no matter what varying stakes you place.

2. If $\alpha_1, \alpha_2, \cdots$ is a nonnegative bounded sequence and $Y_0, Y_1, Y_2, \cdots$ is a supermartingale then $Z_0, Z_1, Z_2, \cdots$ is also a supermartingale. This means that an unfavorable game turns into an unfavorable one no matter what varying stakes you place.

3. If $\alpha_1, \alpha_2, \cdots$ is a nonnegative bounded sequence and $Y_0, Y_1, Y_2, \cdots$ is a submartingale then $Z_0, Z_1, Z_2, \cdots$ is also a submartingale. This means that an favorable game turns into an favorable one no matter what varying stakes you place.

A random variable τ with values in the set $\{1, 2, 3, \cdots\} \cup \{\infty\}$ is called a *stopping time* with respect to filtration $\mathcal{F}_n$ if for each of $n = 1, 2, \cdots$ we have $\{\tau = n\} \in \mathcal{F}_n$.

The following two conditions are in fact equivalent:

1. $\{\tau \leq n\} \in \mathcal{F}_n$ for each $n = 1, 2, \cdots$

2. $\{\tau = n\} \in \mathcal{F}_n$ for each $n = 1, 2, \cdots$

Proof:

$$\begin{aligned} &\{\tau \leq n\} \in \mathcal{F}_n \\ \Longrightarrow &\{\tau \leq n-1\} \in \mathcal{F}_{n-1} \subset \mathcal{F}_n \end{aligned}$$

We can express $\{\tau = n\} = \{\tau \leq n\} - \{\tau \leq n-1\} \in \mathcal{F}_n$, since complements of F_n-measurable sets are F_n-measurable.
Similarly,

$$\begin{aligned} &\{\tau = n\} \in \mathcal{F}_n \\ \implies &\{\tau = 1\} \cup \{\tau = 2\} \cup \cdots \cup \{\tau = n\} \end{aligned}$$

Note that $\{\tau = i\} \in \mathcal{F}_i \subset \mathcal{F}_n$ for $i = 1, 2, \cdots n$. Then $\{\tau = n\} \in \mathcal{F}_n$ since unions of F_n-measurable sets are F_n-measurable.

Now consider the following example: toss a coin repeatedly and win or lose a dollar depending on the outcome. You start with an initial amount of five dollars. Play until you either end up with ten dollars of you lose all that you have.

Denote Y_n to be the amount of money you have at step n. Then

$$\tau = \min\{n | Y_n = 10 \text{ or } 0\}$$

Then τ is a stopping time with respect to the filtration $\mathcal{F}_n = \sigma(X_1, \cdots, X_n)$ since

$$\{\tau = n\} = \{0 < Y_1 < 10\} \cap \cdots \cap \{0 < Y_{n-1} < 10\} \cap \{Y_n = 10 \text{ or } 0\}$$

and each of the events above is $\mathcal{F}_n$-measurable. Hence, the event $\{\tau = n\}$ is also $\mathcal{F}_n$-measurable.

Let Y_n be a sequence of random variables adapted to a filtration $\mathcal{F}_n$ and let $B \subset \mathbb{R}$ be a Borel set. Show that the time of the first entry of Y_n into B:

$$\tau = \min\{n | Y_n \in B\}$$

is a stopping time.

Solution: Note that for any n,

$$\{\tau = n\} = \{Y_1 \notin B\} \cap \cdots \cap \{Y_{n-1} \notin B\} \cap \{Y_n \in B\}$$

where B is a Borel set.

This implies that each of the components of the intersection above are $\mathcal{F}_n$-measurable.

$$\begin{aligned} \{Y_1 \in B\}^C &\in \mathcal{F}_1 \subset \mathcal{F}_n \\ \{Y_2 \in B\}^C &\in \mathcal{F}_2 \subset \mathcal{F}_n \\ &\vdots \\ \{Y_{n-1} \in B\}^C &\in \mathcal{F}_{n-1} \subset \mathcal{F}_n \end{aligned}$$

Hence, $\{\tau = n\} \in \mathcal{F}_n$

Now suppose Y_n is a sequence of random variables adapted to filtration $\mathcal{F}_n$ and let τ be the stopping time with respect to the same filtration. More specifically, let Y_n denote your winnings/losses after n rounds of a game.

This implies that your total winnings are Y_τ is you quit after τ rounds of the game.

Hence, your total winnings after n rounds is $Y_{\min(\tau,n)} =: Y_{\tau \wedge n}$. We will define this to be the *sequence stopped at* τ

$$Y_n{}^{\tau}(\omega) = Y_{\tau(\omega)\wedge n}(\omega)$$

Let τ be a stopping time. Then

1. If Y_n is a martingale, then so is $Y_{\tau\wedge n}$
2. If Y_n is a supermartingale, then so is $Y_{\tau\wedge n}$
3. If Y_n is a submartingale, then so is $Y_{\tau\wedge n}$

We will consider one specific strategy in game play. Suppose a coin, not necessarily fair, is flipped repeatedly. Denote the sequence of outcomes by $X_1, X_2, \cdots$ where

$$X_i = \begin{cases} +1 \text{ if Heads} \\ -1 \text{ if Tails} \end{cases}$$

Consider the following gambling strategy: in the first turn, you bet one dollar on Heads. If you win, you quit. If you lose, you double your betting amount and play again. You repeat until you win and quit at some point, or you lose all of your money.

Mathematically, the strategy can be defined as follows:

$$\alpha_n = \begin{cases} 2^{n-1} \text{ if } X_1 = \cdots = X_{n-1} = \text{Tails} \\ 0 \text{ otherwise} \end{cases}$$

Then we put

$$Z_n = X_1 + 2X_2 + \cdots + 2^{n-1}X_n$$

and the stopping time

$$\tau = \min\{n : X_n = \text{ Heads}\}$$

With this setting, $Z_{\tau\wedge n}$ will be your total winnings after n rounds. Because it is a martingale, we call this strategy the *'martingale'*.

1. Show that if X_n is a sequence of random variables adapted to filtration $\mathcal{F}_n$, then so is the sequence $X_{\tau\wedge n}$.
2. Show that if a gambler plays 'the martingale strategy', then his expected loss just before the final win is infinite:

$$E(Z_{\tau-1}) = -\infty$$

Optional Stopping Theorem
The *optional stopping theorem* (OST) states that if Y_n is a martingale and τ is a stopping time with respect to a filtration $\mathcal{F}_n$ such that the following conditions hold:

1. $\tau < \infty$ almost surely
2. Y_τ is an integrable random variable
3. $E(Y_n) \to 0$ as $n \to \infty$ for all n s.t. $n < \tau$

then

$$E(Y_\tau) = E(Y_1)$$

For $a, b > 0$, let τ be defined by

$$\tau = \min\{n : Y_n = a \text{ or } Y_n = -b\}$$

for a symmetric random walk Y_n. Then it is easy to see that τ satisfies all the above conditions for the OST.

Let p_a be the probability that Y_n reaches a before $-b$. Then

$$E(Y_\tau) = p_a * a + (1 - p_a) * (-b)$$

and by the OST, we have

$$\begin{aligned} E(Y_\tau) &= E(Y_1) \\ &= 0 \end{aligned}$$

because there is equal probability of going up one or going down one, so the expected value is zero. Hence:

$$\begin{aligned} 0 &= p_a * a + (1 - p_a) * (-b) \\ p_a * (a + b) &= b \\ p_a &= \frac{b}{a + b} \end{aligned}$$

Recall from a previous problem that $Z_n := {Y_n}^2 - n$ is a martingale. Then by OST,

$$\begin{aligned} E(Z_\tau) &= E(Z_1) \\ &= {Y_1}^2 - 1 \\ &= \frac{1}{2}(1)^2 + \frac{1}{2}(-1)^2 - 1 \\ &= 0 \end{aligned}$$

We have

$$\begin{aligned} 0 &= E(Z_\tau) \\ &= E({Y_\tau}^2 - \tau) \\ &= E({Y_\tau}^2) - E(\tau) \\ &= a^2 * p_a + b^2 * (1 - p_a) - E(\tau) \end{aligned}$$

and we get that

$$\begin{aligned} E(\tau) &= a^2 * p_a + b^2 * (1 - p_a) \\ &= (a^2 - b^2) * p_a + b^2 \\ &= (a^2 - b^2) * \frac{b}{a + b} + b^2 \\ &= (a - b) * b + b^2 \\ &= ab \end{aligned}$$

Chapter 4: Problems and Exercises

This section contains various problems taken from my own EE 126 class.

1. The following questions pertain to a finite state Markov Chain that consists of a single periodic class (period > 1). Answer the following questions:

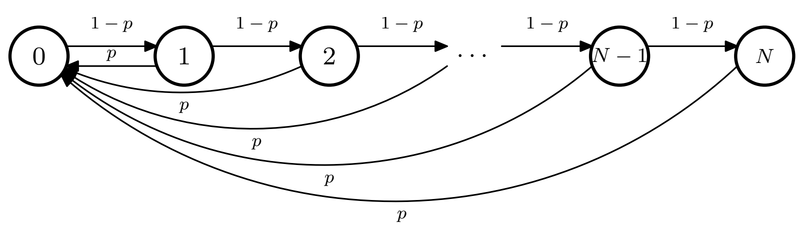
   a. Do all states have to be recurrent?
   Solution: Yes. Because the Markov chain is finite, there must be at least one recurrent state. But there is only a single class, which means that all states must be recurrent.

   b. Is it possible for any of the states to have self-transitions?
   Solution: No. If the chain had self transitions and consisted of a single periodic class, the period of the chain would be exactly 1. However, the period of the chain is given to be strictly greater than 1, a contradiction.

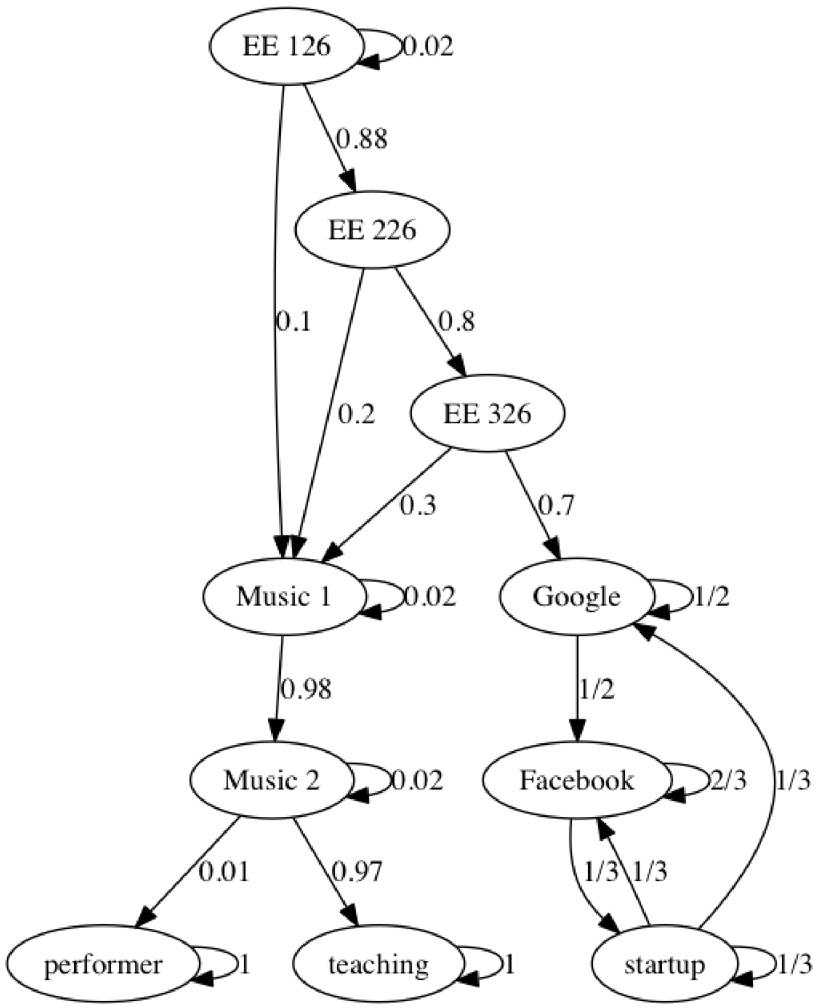
   c. What does the solution to the balance equations tell you about the states of the Markov chain?
   Solution: It tells us the average fraction of time the Markov chains spends in each state.

2. At time $n = 0, 1, 2, \cdots$, a number Y_n of particles is injected into a chamber, where $(Y_n, n \geq 0)$ are independent Poisson random variables with parameter λ. The lifetimes of particles are independent and geometric with parameter p. Let X_n be the number of particles in the chamber at time n. Show that $(X_n, n \geq 0)$ is a Markov chain, find its transition probabilities and its stationary distribution.

3. **Man Climbing Ladder** A man tries to go up a ladder that has N rungs. Every step he takes, he has a probability p of dropping back to the ground and he goes up one rung otherwise. Calculate the average time he takes to reach the top.
Solution: Model the problem as a Markov chain:



4. Modeled below is a Markov chain of the life of the average EE 126 student:



Let S_i denote the state at time i.

(a) Identify both recurrent classes and transient states.
Solution: The recurrent classes are $\{\{4\}, \{5\}, \{8, 9, 10\}, \}$ and the rest of the states are transient states.

(b) If you start in EE 126, what is the probability you end up as a performer? A music teacher? An engineer at a tech company (Google or Facebook or a startup)?
Solution: $\Pr(\text{engineer}) = \sum_{n=0}^{\infty}(0.02)^n(0.88)(0.8)(0.7) = \frac{1}{1-0.02} * (0.88)(0.8)(0.7) \approx 0.503$.
$\Pr(\text{performer}) = (1 - 0.503)\frac{0.01}{0.01+0.97} = 0.005$.
$\Pr(\text{teacher}) = (1 - 0.503)\frac{0.97}{0.01+0.97} = 0.492$.

(c) What is $\lim_{n\to\infty} P(S_n = \text{Google}|S_0 = \text{EE 126})$?
Solution: First label Google as A, Facebook as B, and the startup as C. We set up the

following equations:

$$\frac{1}{2}\pi_A = \frac{1}{3}\pi_C$$
$$\frac{1}{3}\pi_B = \frac{2}{3}\pi_C$$
$$\pi_A + \pi_B + \pi_C = 1$$

Solve to obtain $\pi_A = \frac{2}{11}, \pi_B = \frac{6}{11}$, and $\pi_C = \frac{3}{11}$. Hence $P(S_n = \text{Google}|S_0 = \text{EE 126}) = \frac{2}{11}$.

(d) After leaving Google, what is the mean time of return? (Rest are exercises)

(e) What is $Pr(\text{become engineer at a tech company}|\text{took EE 226})$?

(f) What is the expected time until you start working for Google for the second time, given you become an engineer at a tech company?

5. Consider a continuous time Markov chain with states $\{1, 2, 3, 4\}$ and rate matrix.

$$\begin{bmatrix} -3 & 1 & 1 & 1 \\ 0 & -3 & 2 & 1 \\ 1 & 2 & -4 & 1 \\ 0 & 0 & 1 & -1 \end{bmatrix}$$

 a. Find the stationary distribution π.
 Solution: Solve the equation $\pi Q = 0$ to get $\pi = \begin{bmatrix} \frac{3}{38} & \frac{7}{38} & \frac{9}{38} & \frac{1}{2} \end{bmatrix}$.

 b. Suppose the chain starts in state 1. What is the expected amount of time until it changes state from the first time?
 Solution: Since the distribution of the time spent in state 1 is exponential with rate 3, the expected amount of time spent there is $\frac{1}{3}$.

 c. Assume the chain starts in state 1. What is the expected amount of time until the chain is in state 4?

6. **Machine Maintenance** (Exercise) Consider two machines that are maintained by a single repairman. Machine i functions for an exponentially distributed time with rate λ_i before it fails. The repair times for each unit are exponential with rate μ_i. They are repaired in the order they fail.

 a. Formulate a Markov chain model for this situation. What are the states?

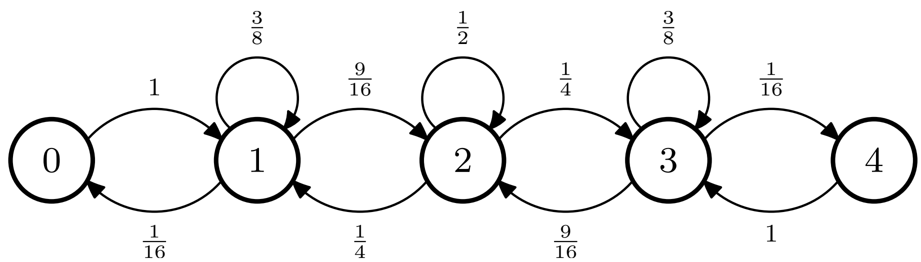
 b. Suppose that $\lambda_1 = 1, \mu_1 = 2, \lambda_2 = 3, \mu_2 = 4$. Find the stationary distribution.

 c. Repeat the problem under the setup where machine 1 now has more importance over machine 2, so the repairman will always service 1 if it's broken.

7. Customers arrive in a shop according to a Poisson process of rate λ. They form a single queue. There are two servers, labeled 1 and 2, with server i requiring an exponentially distributed time of parameter μ_1 to serve any given customer. The service times are independent and independent of the arrival process. The customer at the head of the queue is served by the first idle server; when both are idle, the customer is equally likely to choose either. Describe the Markov process of the queue length and find its stationary distribution. (Exercise)

8. Eight balls, four red and four blue, are distributed in two urns in such a way that each contains four balls. At each step, we draw one ball from each urn at random and place the ball drawn from the first urn into the second, and conversely with the ball from the second urn.

   a. Define a set of states that distinguishes among various color combinations of balls in the two urns. What is the smallest number of states necessary? Draw the appropriate transition diagram.
   Solution: Define the states to be the number of blue balls in the first urn. Then there are five states: 0, 1, 2, 3, and 4. We have the following state transition diagram below:



   b. How many states would you need to distinguish among various color combinations if in addition to the red and blue balls, there are four white balls and the resultant twelve balls are distributed so that there are six balls in each urn?
   Solution: Now we only need to consider the various combinations of two colors instead of one color in the first urn. Choose the two colors to be red and blue. List out all the possible combinations:

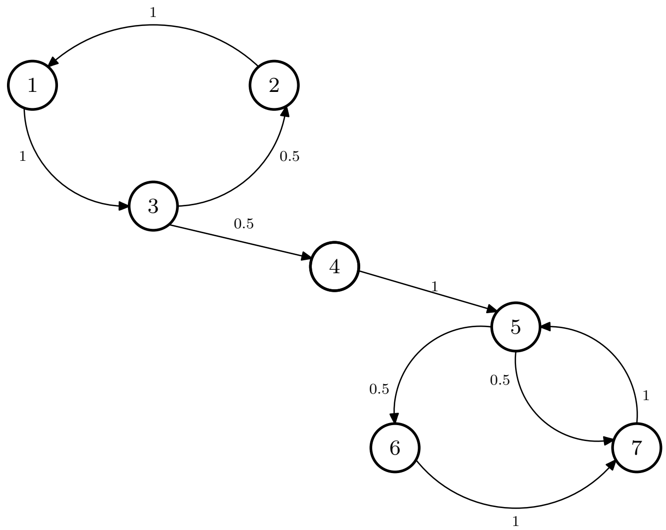
   0 red and 2-4 blue $\rightarrow$ 3 states
   1 red and 1-4 blue $\rightarrow$ 4 states
   2 red and 0-4 blue $\rightarrow$ 5 states
   3 red and 0-3 blue $\rightarrow$ 4 states
   4 red and 0-2 blue $\rightarrow$ 3 states

   Hence, we have a total of 19 states in our new chain.

9. A discrete time Markov chain with seven states has the following transition probabilities:

$$p_{ij} = \begin{cases} 0.5 \text{ if } (i,j) = (3,2),(3,4),(5,6), \text{ and } (5,7) \\ 1 \text{ if } (i,j) = (1,3),(2,1),(4,5),(6,7), \text{ and } (7,5) \\ 0 \text{ otherwise} \end{cases}$$

Let X_k be the state of the Markov process at time k. For what values of n is the probability of reaching state 5 from state 1 in n steps a positive quantity, i.e, $Pr(X_n = 5|X_0 = 1) > 0$?
Solution: The Markov chain diagram is as follows:



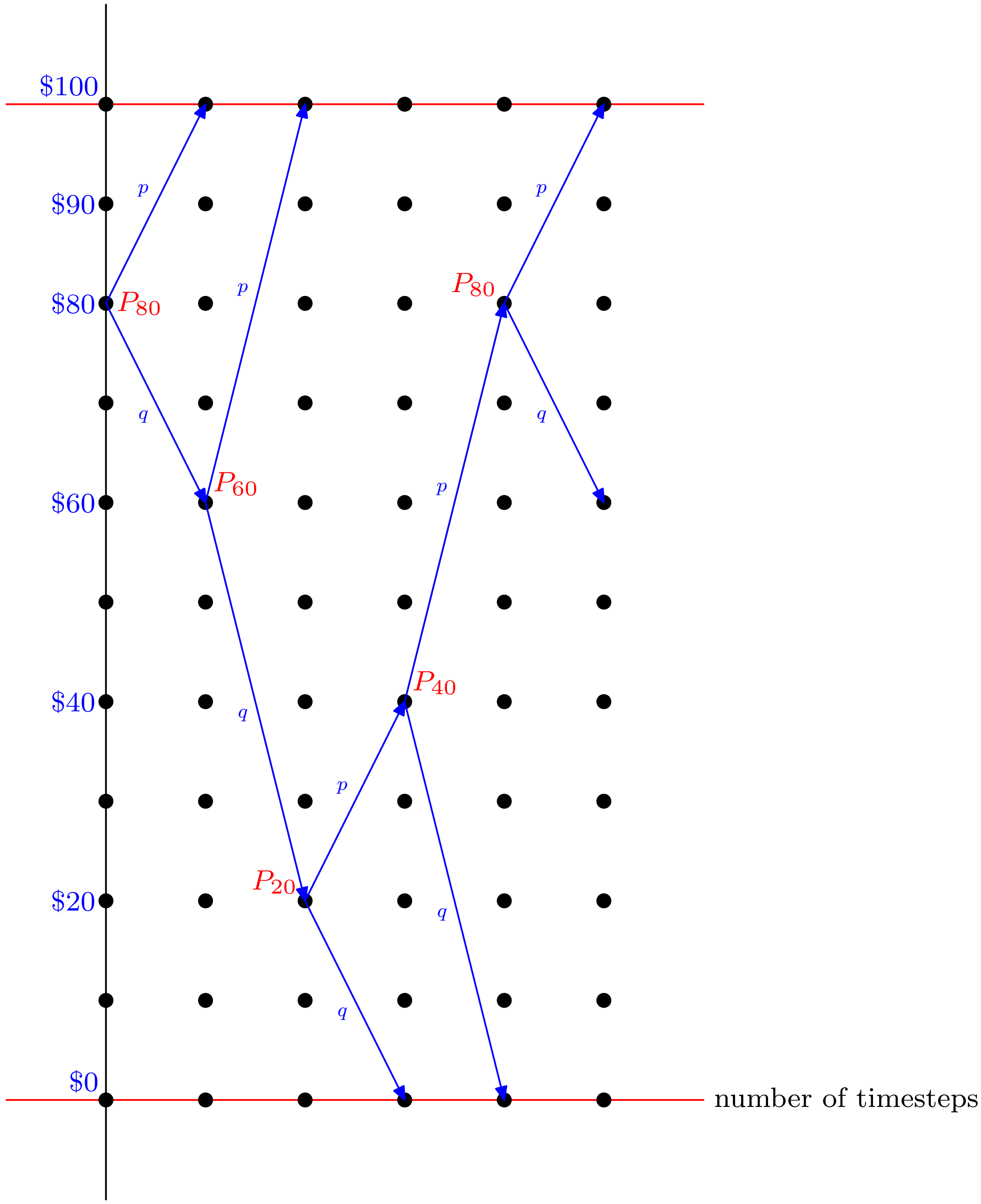
10. There is only one cashier at a certain cafe. Due to limited space, she allows only M customers to line before her at any time. If a customer finds there are M customers there including the one being served at the moment, they will leave the cafe immediately.

    At every minute, exactly one of the following occurs:

    * one new customer arrives with probability p
    * one existing customer leaver with probability kq, where k is the number of customers in the cafe
    * no new customer arrives and no existing customer leaves with probability $1 - p - kq$, if there is at least one customer in the cafe, and with probability $1 - p$ otherwise.

    After the cafe has been open for a long time, you walk in. Calculate how many customers you expect to see in line.

    Solution: Note that this problem can be modeled as a birth-death process. So we have the following Markov chain diagram.



The local balance equations are given as follows:

$$\pi_i * p = \pi_{i+1} * q \qquad\qquad i = 0, \cdots, M-1$$

Then we obtain $\pi_{i+1} = \frac{r}{i+1}\pi_i$ where $r = \frac{p}{q}$. In terms of π_0, we get

$$\pi = \frac{r^i}{i!}\pi_0 \qquad\qquad i = 0, \cdots, M-1$$

Recall that all probabilities must be normalized: $1 = \pi_0 + \pi_1 + \cdots \pi_M$. Then we obtain:

$$1 = \pi_0(1 + \tfrac{r}{1!} + \tfrac{r^2}{2!} + \tfrac{r^3}{3!} + \cdots + \tfrac{r^M}{M!})$$
$$\implies \pi_0 = \left(\sum_{k=0}^{M} \tfrac{r^k}{k!}\right)^{-1}$$

Hence, the steady state probabilities are

$$\pi_i = \tfrac{r^i}{i!}\left(\sum_{k=0}^{M} \tfrac{r^k}{k!}\right)^{-1}$$

And the average number of customers in the cafe is given by

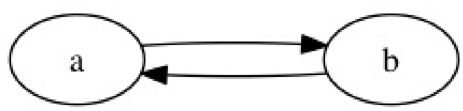
$$N = \sum_{k=0}^{M} k * \pi_k = r * \left(\sum_{k=0}^{M-1} \tfrac{r^k}{k!}\right)\left(\sum_{k=0}^{M} \tfrac{r^k}{k!}\right)^{-1}$$

11. For each statement, determine if it is true or false, along with proper justification.

    a. Every finite Markov chain has at least one recurrent state.
    Solution: True. Suppose we had a finite state Markov chain with no recurrent states (they are all transient states). This means that the time spent in each state is finite. So in the long run, we are in none of the states of the finite chain.

    b. For an irreducible Markov chain with countably infinite states, all states must be transient.
    Solution: False. One counterexample is the typical birth-death process. Enumerate the states $\cdots, -2, -1, 0, 1, 2, \cdots$. Then we have an irreducible Markov chain with countably infinite states. We also have that for every pair of states i, j, i is accessible from j and j is accessible from i. This implies that every state in the chain is recurrent.

    c. If a one-step Markov chain (with transition probabilities given by a matrix P) is irreducible, then the two step Markov chain (with transition probabilities given by P^2) is also irreducible.
    Solution: False. Consider the two state Markov chain as follows:

    Note that its P is irreducible: from either state, we can reach any state in finite time. However, P^2 is not irreducible: two steps taken from state a will lead you directly back to state a again and the same holds for b.

    d. Suppose a state i in a Markov chain is transient. If the chain starts in state i, then the $\{$number of returns to $i\} + 1$ is a geometric random variable.
    Solution: True. Each visit to i has some probability of never returning. Hence, it is geometric.

    e. A finite state irreducible Markov chain has a state i with period 2. It is possible for a different state j to have period 3.
    Solution: False. Every state in a finite state irreducible Markov chain must have the same period. For every state i, the *gcd* of the lengths of all paths from i back to i must be 2. Therefore, no such path can be of length 3.

f. A finite state irreducible Markov chain has a transition matrix P. If a distribution π satisfies $\pi(i)P(i,j) = \pi(j)P(j,i)$ for all states i, j, then π also satisfies $\pi P = \pi$.
Solution: True. Expand the matrix form. The ith element is given by $\sum_j \pi(j)P(j,i)$. Substituting in the given condition, we get that this is equal to $\sum_i \pi(i)P(i,j) = \pi(i)$ because $\sum_i P(i,j) = 1$. Thus, π is a solution to $\pi P = \pi$.

g. A finite state irreducible Markov chain has a symmetric transition matrix P. Then its stationary distribution must be uniform.
Solution: True. For symmetric matrix P and uniform distribution π, we satisfy the condition presented in part f of this problem. Thus the uniform distribution satisfies $\pi P = \pi$ and is a stationary distribution.

12. **Chess Random Walk** If we represent the chessboard as $\{(i,j) : 1 \leq i, j \leq 8\}$ then a knight can move from any (i,j) to any of the eight squares $(i \pm 1, j \pm 2), (i \pm 2, j \pm 1)$, provided they lie on the chessboard. Let X_n be the position of the knight if we pick one of the legal moves at random.

   a. Does this Markov chain have a stationary distribution? If so, find it and the expected time to return to state $(1,1)$ assuming the random walk starts from time $(1,1)$.
   Solution: We have a total of 64 states in our Markov chain. Two states have an edge between them if a knight can legally move from one state to the other.
   For each state (i,j), let $d(i,j)$ denote the number of edges that are outgoing from the state, i.e, the degree of the state. Since the knight moves randomly, transition probabilities are then $\frac{1}{d(i,j)}$.
   Because we are dealing with an irreducible chain, we have a unique stationary distribution $\pi(i,j)$. To find a more precise expression of these quantities, we shall define the following indicator random variable:

$$X_{(i,j)}(a,b) = \begin{cases} 1 \text{ if a knight at } (i,j) \text{ can reach } (a,b) \\ 0 \text{ else} \end{cases}$$

   Then

$$\begin{aligned} d(i,j) &= \sum_{(a,b)} X_{(i,j)}(a,b) \\ &= \sum_{(a,b)} d(a,b) * \frac{X_{(i,j)}(a,b)}{d(a,b)} \\ &= d(a,b) \sum_{(a,b)} \frac{X_{(i,j)}(a,b)}{d(a,b)} \\ &= d(a,b) \sum_{(a,b)} p_{(i,j)}(a,b) \end{aligned}$$

   where $p_{(i,j)}(a,b)$ is the transition probability from (i,j) to (a,b).

After normalizing, we get that

$$\begin{aligned}\pi(i,j) &= \frac{d(i,j)}{\sum\limits_{(a,b)} d(a,b)} \\ &= \frac{2}{336}\end{aligned}$$

since there are only two outgoing edges from state $(1,1)$ and the sum of all the degrees of all states is 336.

b. Repeat the question for a Queen's Random Walk. (A queen can move any number of squares horizontally, vertically, or diagonally).

13. **Gambler's Ruin and Winning Games** Bob, who has 15 pennies, and Charlie, who has 10 pennies, decide to play a game. They flip a fair coin. If the coin is Heads, Bob gets one of Charlie's pennies. If the coin is Tails, Charlie gets one of Bob's pennies. They quit when one of them has all the pennies.

    a. Suppose the coin is fair. Then what is the chance that Bob wins the game?
    Solution: It would be easier for Bob to win all of Charlie's pennies because Charlie has less pennies than Bob. Following the intuition behind random walks, the probability that Bob gets all the pennies before Charlie does is $\frac{15}{10+15} = \frac{15}{25} = \frac{3}{5}$.

    b. Suppose again the coin is fair. What is the expected number of flips until the game ends?

14. Consider the problem where a sole gambler is playing games. She wins each game with probability p. In each of the following cases, determine the expected total number of wins.

    (a) The gambler will first play n games. If she wins X of these games, then she will play X more games, then stop.
    Solution: $E(n+X) = n + E(X) = n + np$

    (b) The gambler will first play until she wins. If it takes her Y games to reach her first win, she will play Y more games, then stop.
    Solution: $E(2Y) = 2E(Y) = \frac{2}{p}$

15. If we have a deck of 52 cards, numbered $\{1, 2, 3, \cdots 52\}$. Then its state can be described by the sequence of numbers we see as we examine the deck from top down. Consider the following *shuffling procedure*: pick a card at random from the deck (including the top card) and place it at the top.

    a. Model the shuffling as a Markov chain. Explain clearly what the states are, how many states there are, what the transition probabilities are like, and determine if the chain is aperiodic or irreducible.
    Solution: This shuffling can be modeled as a Markov chain with 52! states, one for each possible permutation of cards. For each state $(i_1, i_2, \cdots i_{52})$, there are 52 transitions to states of the form $(i_k, i_1, i_2, \cdots, i_{k-1}, i_{k+1}, \cdots, i_{52})$. The transition probabilities are $\frac{1}{52}$ because any card is equally likely to be picked.
    The chain is aperiodic because each state has a self-loop. It is also irreducible because any state $(j_1, \cdots, j_{52})$ can be reached from any state $(i_1, \cdots, i_{52})$ by picking out the cards $(i_1, \cdots, i_{52})$ in reverse order.

b. Show that if we repeatedly apply this algorithm, then in the limit, the deck is perfectly shuffled in the sense that all 52! possibilities are equally likely. This is equivalent to saying that the stationary distribution of the Markov chain is uniform.
Solution: To prove this, we must show that the uniform distribution is the stationary distribution. We have

$$\tfrac{1}{52!} = \pi(x) = \sum_{b \in B} \pi(b) P(b, x) = \sum_{b \in B} \tfrac{1}{52!} \tfrac{1}{52} = \tfrac{1}{52!}$$

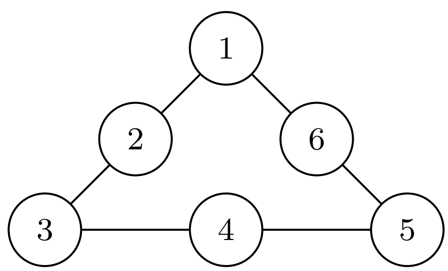
since every transition probability is $\frac{1}{52}$. Thus, the uniform distribution $\{\frac{1}{52}, \frac{1}{52}, \cdots \frac{1}{52}\}$ satisfies $\pi P = P$ and so is the stationary distribution.

a. Argue that the queue length is a Markov chain. Draw the transition diagram of the Markov chain.

b. Prove that for all finite values of λ and μ, the Markov chain is positive recurrent and find the invariant distribution.

c. Now consider a different setup. The continuous time queue still has Poisson arrivals with rate λ, but it instead has just two servers that work in parallel. Where there are at least two customers in the queue, two are being served. When there is only one customer, only one server is active. The service times are iid $Exp(\mu)$. Find the minimum value of μ so that the queue is positive recurrent and solve the balance equations.

16. Consider a discrete time Markov process $\{X_k : k > 0\}$ with state space $\{1, 2, 3, 4, 5, 6\}$. Suppose the states are arranged in the triangle shown:



Given $X_k = i$, the next state X_{k+1} is one of the two neighbors of i selected with probability 0.5 each. Suppose $P(X_0 = 1) = 1$.

a. Let $\tau_B = min\{k : X_k \in \{3, 4, 5\}\}$ (the first time the base of the triangle is reached). Find $E[\tau_B]$.
Solution: Set states 3, 4, and 5 to be absorbing states and obtain the following equations, where a_i represents the number of steps needed to take to get absorbed, starting from state i.

$$a_1 = 1 + \frac{1}{2}a_2 + \frac{1}{2}a_6$$
$$a_2 = 1 + \frac{1}{2}a_1 + \frac{1}{2}a_3 a_3 = a_4 = a_5 = 0$$
$$a_6 = a_2$$

Our desired quantity is then a_1, and solving the system above yields $a_1 = 4$.

b. Let $\tau_3 = min\{k : X_k = 3\}$. Find $E[\tau_3]$.
Solution: Set state 3 to be the absorbing state and obtain the following equations in a similar fashion to those of part a.

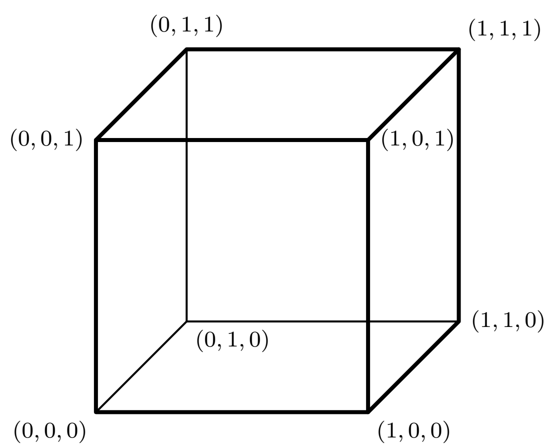
$$a_1 = 1 + \frac{1}{2}a_2 + \frac{1}{2}a_6 a_2 = 1 + \frac{1}{2}a_1 a_3 = 0 a_4 = a_2 a_5 = a_1 a_6 = 1 + \frac{1}{2}a_5 + \frac{1}{2}a_1$$

Again, we desire a_1, and solving the above set of equations yields $a_1 = 8$.

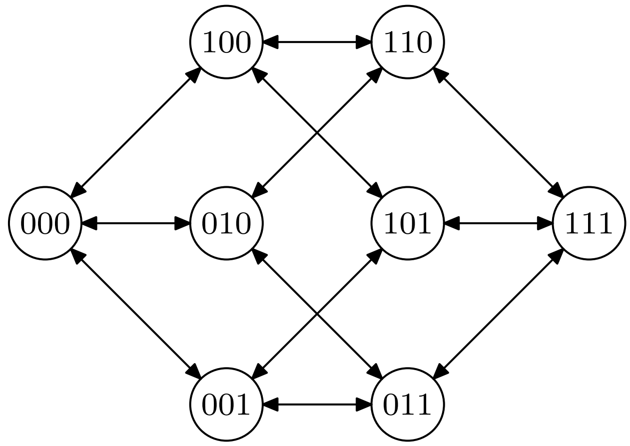
c. Let τ_C be the first time $k > 1$ such that both states 3 and 5 have been visited by time k. Find $E[\tau_C]$. (Use the results of parts a and b as well as symmetry).
Solution: Note that the diagram is symmetric. Furthermore, we have $E[\tau_C] = E[\tau_3] + E[\tau_B] = 8 + 4 = 12$.

d. Let τ_R denote the first time $k > TC$ such that $X_k = 1$. That is, τ_R is the first time the process returns to vertex 1 of the triangle after reaching both of the other vertices. Find $E[\tau_R]$. (Use the results of parts b and c as well as symmetry).
Solution: Note that the diagram is symmetric. To get from $3 \to 5 \to 1$ or $5 \to 3 \to 1$, just add the solution of part b to that of part c.

17. **Fly on a Cube** Consider a cube with vertices 000, 001, 010, 100, 110, 101, 011, 111. Suppose a fly walks along the edges of the cube from vertex to vertex and for any integer $t \geq 0$, let X_t denote which vertex the fly is at time t. Assume $X = \{X_t, t \geq 0\}$ is a discrete time Markov process, such that given X_t, the next state X_{t+1} is equally likely to be any one of the three vertices neighboring X_t.



Suppose the fly begins at vertex 000 at time zero. Let T be the first time that X returns to vertex 000 after time 0. Find $E[T]$.
Solution: It may be easier to consider the following flattened-out cube instead:



Obtain a set of equations from the above graph. Let μ_{abc} denote the expected time at which we reach state abc from state 000. Denote μ^*_{000} to be the first time we reach state 000 after time 0. This is the quantity we desire.

$\mu_{000} = 0$ since the first time we reach state 000 overall if we start from state 000 is at time 0. And by symmetry of the cube with respect to the distance from the starting state 000, we have

$$\mu_{100} = \mu_{010} = \mu_{001}$$
$$\mu_{101} = \mu_{110} = \mu_{011}$$

The simplified set of equations we obtain is:

$$\begin{aligned}
\mu^*_{000} &= 1 + \frac{1}{3}\mu_{001} + \frac{1}{3}\mu_{100} + \frac{1}{3}\mu_{010} \\
&= 1 + \mu_{001} \\
\mu_{001} &= 1 + \frac{1}{3}\mu_{011} + \frac{1}{3}\mu_{101} + \frac{1}{3}\mu_{000} \\
&= 1 + \frac{2}{3}\mu_{011} + \frac{1}{3}\mu_{000} \\
&= 1 + \frac{2}{3}\mu_{011} \\
\mu_{011} &= 1 + \frac{1}{3}\mu_{010} + \frac{1}{3}\mu_{001} + \frac{1}{3}\mu_{111} \\
&= 1 + \frac{2}{3}\mu_{001} + \frac{1}{3}\mu_{111} \\
\mu_{111} &= 1 + \frac{1}{3}\mu_{011} + \frac{1}{3}\mu_{110} + \frac{1}{3}\mu_{101} \\
&= 1 + \mu_{011}
\end{aligned}$$

Solving gives us $\mu^*_{000} = 8$. So it takes an expected time of 8 time steps to return back to 000, once started from 000.

Now consider the case where the fly stays at its current location with probability $\frac{1}{2}$ and that each edge still has equal probability of being chosen by the fly to walk along. Solve the same problem: given that the fly starts at the vertex in the center of the large cube, what is the expected time of return?

Solution: Set up similar equations as above and use the same notation.

Again we have $\mu_{000} = 0$ since the first time we reach state 000 overall if we start from state 000 is at time 0. And the symmetry of the cube still holds, so:

$$\mu_{100} = \mu_{010} = \mu_{001}$$
$$\mu_{101} = \mu_{110} = \mu_{011}$$

The simplified set of equations we obtain is:

$$\begin{aligned}
\mu^*_{000} &= \frac{1}{2}(1) + \frac{1}{6}(1+\mu_{001}) + \frac{1}{6}(1+\mu_{100}) + \frac{1}{6}(1+\mu_{010}) \\
&= 1 + \frac{1}{2}\mu_{001} \\
\mu_{001} &= 1 + \frac{1}{2}\mu_{001} + \frac{1}{6}\mu_{011} + \frac{1}{6}\mu_{101} + \frac{1}{6}\mu_{000} \\
&= 1 + \frac{1}{2}\mu_{001} + \frac{1}{3}\mu_{011} + \frac{1}{3}\mu_{000} \\
&= 1 + \frac{1}{2}\mu_{001} + \frac{1}{3}\mu_{011} \\
\mu_{011} &= 1 + \frac{1}{2}\mu_{011} + \frac{1}{6}\mu_{010} + \frac{1}{6}\mu_{001} + \frac{1}{6}\mu_{111} \\
&= 1 + \frac{1}{2}\mu_{011} + \frac{1}{3}\mu_{001} + \frac{1}{6}\mu_{111} \\
\mu_{111} &= 1 + \frac{1}{2}\mu_{111} + \frac{1}{6}\mu_{011} + \frac{1}{6}\mu_{110} + \frac{1}{6}\mu_{101} \\
&= 1 + \frac{1}{2}\mu_{111} + \frac{1}{2}\mu_{011}
\end{aligned}$$

Solving gives us $\mu^*_{000} = 8$. So it takes an expected time of 8 time steps to return back to 000, once started from 000. Note that although we have different transition probabilities, the expected number of times to return in the same!

Now consider an extension of the problem where instead of a three-dimensional cube, we have a n-dimensional cube. Use the probabilities in the very first consideration from above, where transition states are reached from any state with probability $\frac{1}{n}$, since there will be n edges coming from a vertex in an n-dimensional cube. We wish to solve the same problem: starting from vertex $(0, 0, \cdots, 0)$, what is the expected number of jumps needed to come back to $(0, 0, \cdots, 0)$?

Solution: As a three-dimensional cube has eight vertices and therefore, eight states in the Markov chain, a n-dimensional cube will have 2^n states.

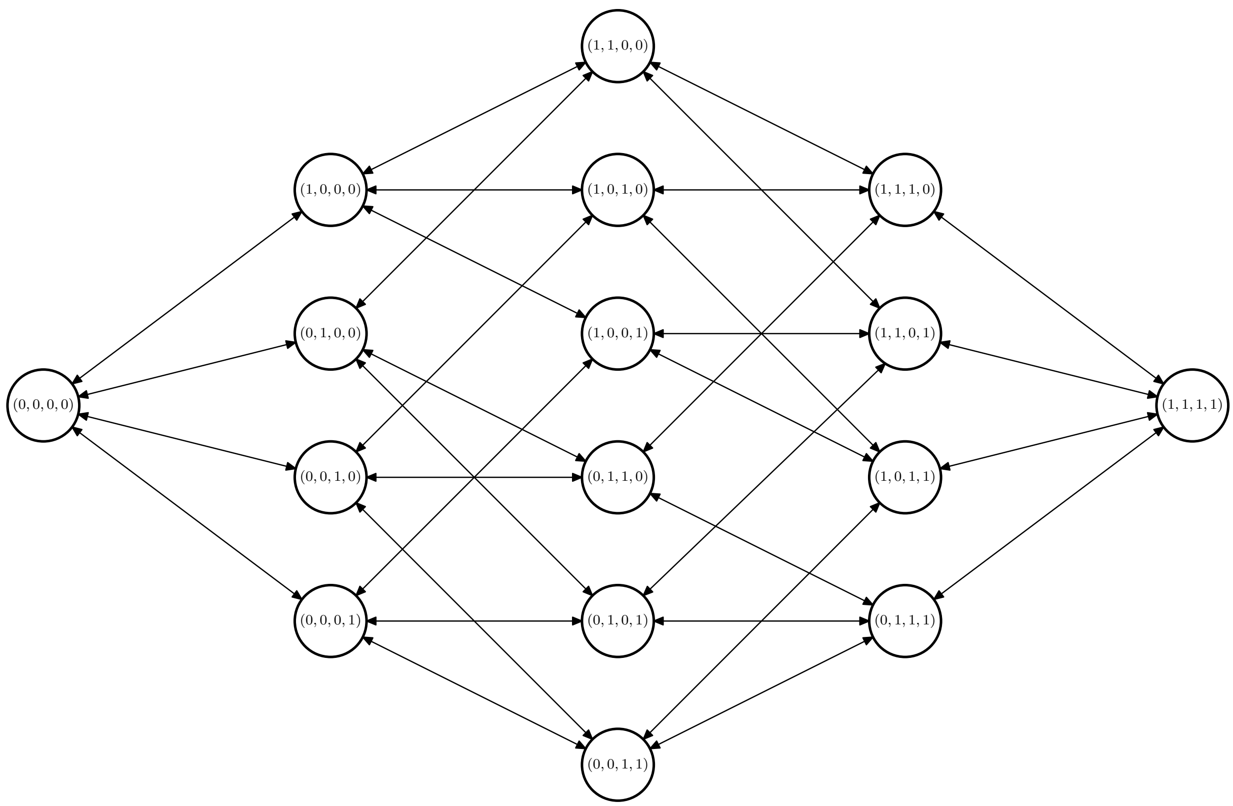
Let us first determine which states can transition to which states. Extending from the three dimensional case, we have that any state $(x_1, x_2, x_3, \cdots x_n)$, $x_i \in \{0, 1\}$, transitions to any other state that has exactly one of the coordinate values offset by ± 1. So for instance $(0, 0, \cdots, 0)$ can transition to any of the n states $(1, 0, 0, \cdots, 0), (0, 1, 0, \cdots 0), (0, 0, 1, \cdots, 0), \cdots, (0, 0, 0, \cdots, 1)$. Furthermore, each state will have n outgoing edges, among which one of them are chosen randomly. So the transition probabilities for each one is $\frac{1}{n}$.

Let us use the notation $\mu(x_1, x_2, \cdots, x_n)$ to denote the mean time it takes to reach $(0, 0, \cdots 0)$ from state $(x_1, x_2, \cdots, x_n)$. Further denote $\mu(0, 0, \cdots 0)^*$ to be the first time we reach state $(0, 0, \cdots 0)$ after time 0. We have by symmetry:

$$\mu(1, 0, \cdots 0) = \mu(0, 1, \cdots 0) = \cdots \mu(0, 0, \cdots 1)$$

Similarly, all the states with k out of n coordinates equal to 1 $(k < n)$ will have the same μ by symmetry.

Before we set up our equations, let us first consider the simpler case where $n = 4$. We have the following graph of transition states:



Note that for any state (x_1, x_2, x_3, x_4) with 1 one, three out of the four transitions are made to a state with 2 ones and one is made to a state with no ones. For any state with 2 ones, two out of four transitions are made to a state with 3 ones and two are made to a state with 1 one. And we established above that by symmetry, all the states with exactly k ones, with

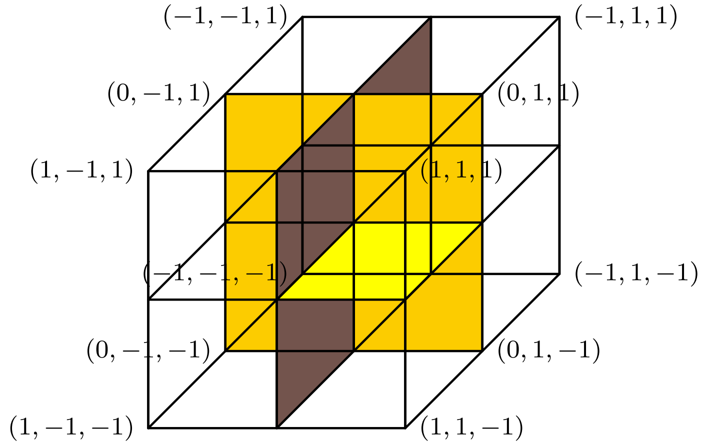
$k \in \{1, 2, 3\}$ share the same mean time μ. Hence our equations are as follows:

$$\begin{aligned}
\mu(0,0,0,0)^* &= 1 + \frac{1}{4}\mu(1,0,0,0) + \frac{1}{4}\mu(0,1,0,0) + \frac{1}{4}\mu(0,0,1,0) + \frac{1}{4}\mu(0,0,0,1) \\
&= 1 + \mu(1,0,0,0) \\
\mu(1,0,0,0) &= 1 + \frac{1}{4}\mu(0,0,0,0) + \frac{1}{4}\mu(1,1,0,0) + \frac{1}{4}\mu(1,0,1,0) + \frac{1}{4}\mu(1,0,0,1) \\
&= 1 + \frac{3}{4}\mu(1,1,0,0) \\
\mu(1,1,0,0) &= 1 + \frac{1}{4}\mu(1,0,0,0) + \frac{1}{4}\mu(0,1,0,0) + \frac{1}{4}\mu(1,1,1,0) + \frac{1}{4}\mu(1,1,0,1) \\
&= 1 + \frac{1}{2}\mu(1,0,0,0) + \frac{1}{2}\mu(1,1,1,0) \\
\mu(1,1,1,0) &= 1 + \frac{1}{4}\mu(0,1,1,0) + \frac{1}{4}\mu(1,0,1,0) + \frac{1}{4}\mu(1,1,0,0) + \frac{1}{4}\mu(1,1,1,1) \\
&= 1 + \frac{3}{4}\mu(1,1,0,0) + \frac{1}{4}\mu(1,1,1,1) \\
\mu(1,1,1,1) &= 1 + \frac{1}{4}\mu(0,1,1,1) + \frac{1}{4}\mu(1,0,1,1) + \frac{1}{4}\mu(1,1,0,1) + \frac{1}{4}\mu(1,1,1,0) \\
&= 1 + \mu(1,1,1,0)
\end{aligned}$$

We can generalize this to a state of n dimensions. For any state $(x_1, x_2, \cdots, x_n)$, with k ones, $n-k$ of the transitions we make are to a state with $k+1$ ones and k of the transitions we make are to a state with $k-1$ ones. Then we can use the same symmetry argument to simplify our set of equations.

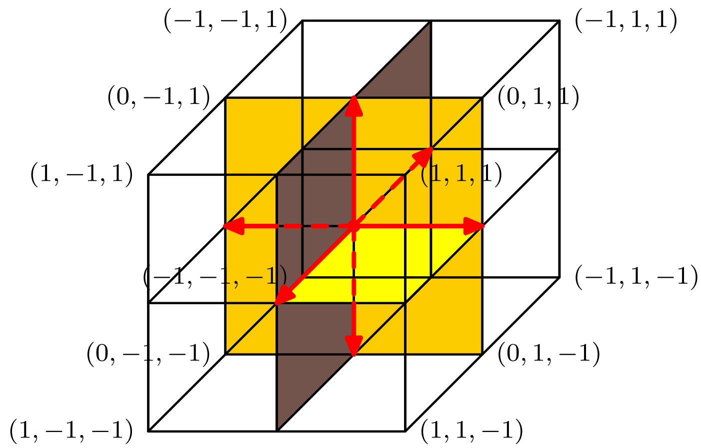
$$\begin{aligned}
\mu(0,0,0,0,\cdots,0)^* &= 1 + \mu(1,0,0,0,\cdots,0) \\
\mu(1,0,0,0,\cdots,0) &= 1 + \frac{n-1}{n}\mu(1,1,0,0,\cdots,0) \\
\mu(1,1,0,0,\cdots,0) &= 1 + \frac{2}{n}\mu(1,0,0,0,\cdots,0) + \frac{n-2}{n}\mu(1,1,1,0,\cdots,0) \\
\mu(1,1,1,0,\cdots,0) &= 1 + \frac{3}{n}\mu(1,1,0,0,\cdots,0) + \frac{n-3}{n}\mu(1,1,1,1,0,\cdots,0) \\
&\vdots \\
\mu(1,1,1,1,\cdots,1) &= 1 + \mu(1,1,1,\cdots,1,0)
\end{aligned}$$

Now consider a separate extension of the problem where we have eight copies of the original three dimensional cube put together into one large cube.



The fly can only walk along the edges of the cubes that are stuck together, and from each vertex, each edge has equal probability of being chosen by the fly to walk along, and it never stays at one vertex for more than one time step. Given that the fly starts at the vertex in the center of the large cube, what is the expected time of return?

Solution: Starting from the center, there are six different directions the fly can walk along



Hence, each direction is taken with probability $\frac{1}{6}$. We get the following equation:

$$\begin{aligned}\mu(0,0,0)^* &= 1 + \frac{1}{6}\mu(1,0,0) + \frac{1}{6}\mu(-1,0,0) + \frac{1}{6}\mu(0,1,0) + \frac{1}{6}\mu(0,-1,0) + \frac{1}{6}\mu(0,0,1) + \frac{1}{6}\mu(0,0,-1)\\ &= 1 + \mu(1,0,0)\end{aligned}$$

where the simplification can be made because of symmetry.

Similarly, starting from the center of a face, for example, vertex $(1, 0, 0)$, there are five vertices we can go towards:

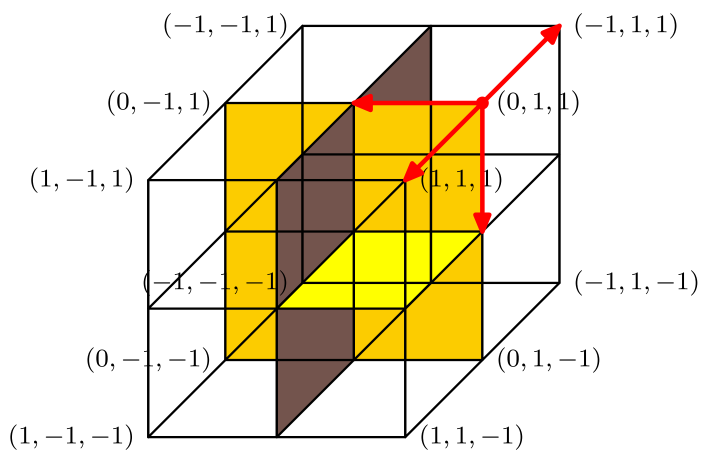
$$\begin{aligned}&(0,0,0)\\&(1,1,0)\\&(1,-1,0)\\&(1,0,1)\\&(1,0,-1)\end{aligned}$$

each of them with probability $\frac{1}{5}$. Our next equation is as follows:

$$\begin{aligned}\mu(1,0,0) &= 1 + \frac{1}{5}\mu(0,0,0) + \frac{1}{5}\mu(1,1,0) + \frac{1}{5}\mu(1,-1,0) + \frac{1}{5}\mu(1,0,1) + \frac{1}{5}\mu(1,0,-1)\\ &= 1 + \frac{4}{5}\mu(1,1,0)\end{aligned}$$

where again the simplification can be made because of symmetry.

Now starting from $(1,1,0)$, there are four directions that can be taken: (the figure below considers the state $(0,1,1)$ instead of $(1,1,0)$ but their μ are the same by symmetry).



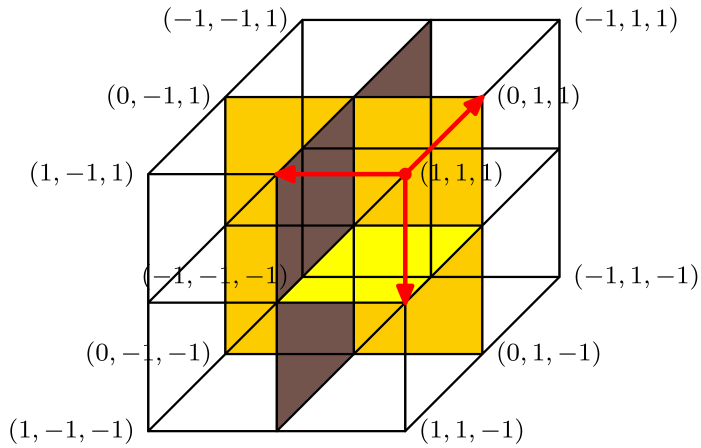
so the transition probabilities are $\frac{1}{4}$ each.

Set up the following equation:

$$\begin{aligned}\mu(1,1,0) &= 1 + \frac{1}{4}\mu(0,1,0) + \frac{1}{4}\mu(1,0,0) + \frac{1}{4}\mu(1,1,1) + \frac{1}{4}\mu(1,1,-1)\\ &= 1 + \frac{1}{2}\mu(1,0,0) + \frac{1}{2}\mu(1,1,1)\end{aligned}$$

where again the simplification can be made because of symmetry.

Finally, we shall cover the corner case of starting at $(1,1,1)$.



Set up the equation similarly:

$$\begin{aligned}\mu(1,1,1) &= 1 + \frac{1}{3}\mu(0,1,1) + \frac{1}{3}\mu(1,0,1) + \frac{1}{3}\mu(1,1,0) \\ &= 1 + \mu(1,1,0)\end{aligned}$$

Set up equations similar to the ones formed for the single cube case.

$$\begin{aligned}\mu(0,0,0)^* &= 1 + \mu(1,0,0) \\ \mu(1,0,0) &= 1 + \frac{4}{5}\mu(1,1,0) \\ \mu(1,1,0) &= 1 + \frac{1}{2}\mu(1,0,0) + \frac{1}{2}\mu(1,1,1) \\ \mu(1,1,1) &= 1 + \mu(1,1,0)\end{aligned}$$

Solving the set of equations, we get that $\mu(1,0,0) = 17, \mu(1,1,0) = 20, \mu(1,1,1) = 21$, and $\mu(0,0,0)^* = 18$. So it takes an expected time of 18 time steps to return back to $(0,0,0)$, once started from $(0,0,0)$.

Now consider the case where the fly stays at its current location with probability $\frac{1}{2}$ and that each edge still has equal probability of being chosen by the fly to walk along. Solve the same problem: given that the fly starts at the vertex in the center of the large cube, what is the expected time of return?

Solution: We now set up the following equations instead.

$$\begin{aligned}\mu(0,0,0)^* &= 1 + \mu(1,0,0) \\ \mu(1,0,0) &= 1 + \frac{1}{2}\mu(1,0,0) + \frac{4}{5}\mu(1,1,0) \\ \mu(1,1,0) &= 1 + \frac{1}{2}\mu(1,1,0) + \frac{1}{2}\mu(1,0,0) + \frac{1}{2}\mu(1,1,1) \\ \mu(1,1,1) &= 1 + \frac{1}{2}\mu(1,1,1) + \mu(1,1,0)\end{aligned}$$

Solving, we get that $\mu(1,0,0) = 34, \mu(1,1,0) = 40, \mu(1,1,1) = 42$, and $\mu(0,0,0)^* = 18$. So once again, it takes an expected time of 18 time steps to return back to 0, once started from 0. Note that the expected times to return are the same, despite the different probabilities!

18. Coin 1 comes up heads with probability 0.6 and coin 2 with probability 0.5. A coin is continually flipped until it comes up tails, at which time that coin is put aside and we start flipping the other one.

    a. What proportion of the flips use coin 1?
    Solution: Construct a two state Markov chain where the state at time n denotes which of the two coins is being flipped at that time n.

    The stationary probabilities satisfy the following equations:

$$\pi_1 = 0.6\pi_1 + 0.5\pi_2$$
$$\pi_2 = 0.5\pi_2 + 0.4\pi_1$$
$$\pi_1 + \pi_2 = 1$$

    Solving them yields $\pi_1 = \frac{5}{9}, \pi_2 = \frac{4}{9}$. Hence, the proportion of coin 1 flips is $\frac{5}{9}$.

    b. If we start the process with coin 1, what is the probability that coin 2 is used on the fifth flip?
    Solution: From the Markov chain figure above, map out all the different trajectories that start with state 1 and end with state 2 in four steps.

$$1 \to 1 \to 1 \to 1 \to 2$$
$$1 \to 1 \to 1 \to 2 \to 2$$
$$1 \to 1 \to 2 \to 2 \to 2$$
$$1 \to 2 \to 2 \to 2 \to 2$$
$$1 \to 1 \to 2 \to 1 \to 2$$
$$1 \to 2 \to 1 \to 1 \to 2$$
$$1 \to 2 \to 2 \to 1 \to 2$$
$$1 \to 2 \to 1 \to 2 \to 2$$

    Compute the probabilities associated with each trajectory. For example, $1 \to 1 \to 1 \to 1 \to 2$ has probability $0.6 * 0.6 * 0.6 * 0.4 = 0.0864$. Repeat the calculation for each of the other trajectories and sum up the results to get 0.444. Hence, the probability that coin 2 is used on the second flip, given that coin 1 is used on the first flip is 0.444.

19. Let Y_n be the sum of n independent rolls of a fair die. Find $\lim_{n\to\infty} P(Y_n \text{ is a multiple of } 13)$.
    Solution: Let X_n denote that value of Y_n modulo 13. Then X_n is a finite state Markov chain with thirteen states enumerated $0, 1, 2, \cdots 12$.

    The state transition matrix is as follows:

$$\begin{bmatrix}
0 & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \\
\frac{1}{6} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \\
\frac{1}{6} & \frac{1}{6} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \\
\frac{1}{6} & \frac{1}{6} & \frac{1}{6} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \\
\frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{6} & \frac{1}{6} \\
\frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{6} \\
\frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{bmatrix}$$

Note that this is a doubly stochastic matrix. So by symmetry, the limiting probabilities are identically equal to $\frac{1}{13}$.

20. **Die Patterns** Toss three dice. Find the probability of getting a sum (the sum of the numbers of two consecutive rolls) of 5 before getting a sum of 7.
Solution: Let A denote the event that we get a sum of 5 before a sum of 7 and p be the probability of A occurring. Let $P(n)$ denote the probability of getting a sum of k.

Using conditional probabilities, we get

$$\begin{aligned}
p &= P(A|\text{sum of } 5)P(5) + P(A|\text{sum of } 7)P(7) + P(A|\text{sum of neither})(1 - P(5) - P(7)) \\
&= 1 * P(5) + 0 * P(7) + p * (1 - P(5) - P(7))
\end{aligned}$$

Solving for p yields:

$$\begin{aligned}
(P(5) + P(7)) * p &= P(5) \\
p &= \frac{P(5)}{P(5) + P(7)}
\end{aligned}$$

And we can manually count the number of ways to obtain a sum of 5 or 7 from three die. Divide each quantity by the total types of sums to get $P(5)$ and $P(7)$.

$$\begin{aligned}
P(5) &= \frac{6}{6^3} \\
P(7) &= \frac{21}{6^3}
\end{aligned}$$

so

$$p = \frac{6}{21} = \frac{2}{7}$$

21. **More Die Patterns** (Exercise) Toss a die until you get a sequence of numbers. Calculate the following:

    a. Expected number of tosses until the last three are 666.

b. Expected number of tosses until the last two are 12.

c. Expected number of tosses until the last six are 123456.

22. **Even More Die Patterns** You roll a die successively until the sum of the last two numbers is 10. Find the average number of times you roll the die. (Exercise)