Math 450 Homework 5
Markov chains to Simulations

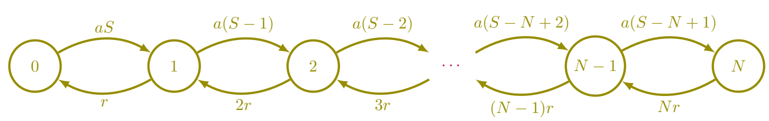
due April 5, 2021

1. If telephone calls arrive at a rate of 1 per minute and last an average of 8 minutes, how many phone lines would you need from a neighborhood to make sure the system is maxed out no more than 1% of the time?

   Erlang's B formula
   $$P_N = \frac{1}{N!}\left(\frac{\lambda}{r}\right)^N \left(\sum_{j=0}^{N} \frac{(\lambda/r)^j}{j!}\right)^{-1}$$
   doesn't have an explicit solution for N, but we can calculate the probability as a function of N and inspect the results. When we do that, we see that 14 lines are congested 1.72 percent of the time, and 15 lines are congested 0.91 percent of the time, so at least 15 lines are needed to meet our desired reliability criteria. If you have studied some numerical methods, you might consider the problem of most quickly performing these calculations.

2. When the number of subscribers is relatively small, Erlang's theory is not a good approximation because as more calls are placed, the number of subscribers who might place a new call decreases. In 1918, Tore Engset published an analysis of this problem.

   (a) Assuming there are S subscribers and N lines with $N < S$, that the hazard rate for each individal subscriber to place a call is a and that the hazard rate for each individual call to terminate is r, and that blocked calls are lost, draw a digraph of the continuous- time Markov chain for the number of lines in use at a given time.

   (b) Construct the forward equation describing the change in each state probability for this "continuous-time Markov chain."
   $$\dot{p}_j = r(j+1)p_{j+1} - (a(S-j)+rj)p_j + a(S-j+1)p_{j-1}$$
   except when $j = 0, N$, in which special cases,
   $$\dot{p}_0 = rp_1 - aSp_0$$
   $$\dot{p}_N = a(S-N+1)p_{N-1} - rNp_N$$
   You could also write this out in matrix form for M, where $\dot{p} = Mp$.

   (c) At steady-state, solve for p_1 in terms of p_0. Then solve for p_2 in terms of p_1, and so on. Find a pattern for the p_j's term.

   $p_1 = \dfrac{aS}{r}p_0$, $p_2 = \left(\dfrac{a}{r}\right)^2 \dfrac{S(S-1)}{2}p_0$, and in general $p_j = \left(\dfrac{a}{r}\right)^j \dfrac{S!}{j!(S-j)!}p_0$,

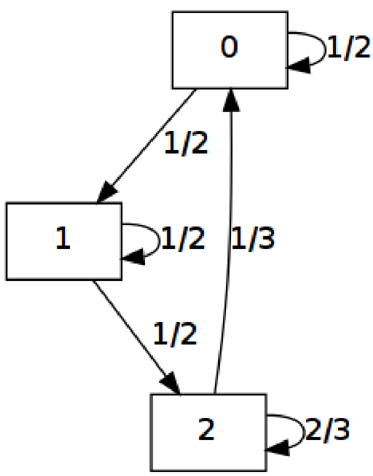
(d) Under what conditions will Engset's solution and Erlang's solution approximately agree?

Engset's solution is a truncated binomial distribution while Erlang's solution is a truncated Poisson distribution, so we expect their solutions to agree when S is very large and N is small, so $S!/(S-j)! \approx S^j$.

* **(warm-up, not for credit)** The following examples are to give you some practice with Markov chain calculations. For each transition matrix, draw a directed graph, with nodes labeled for each state and edges labeled with their transition probabilities. Then, try to find the equilibrium distribution $\tilde{p}(\infty)$ such that $\tilde{p}(\infty) = A\tilde{p}(\infty)$ and discuss the meaning of your result.

(a)

$$A := \begin{bmatrix} \frac{1}{2} & 0 & \frac{1}{3} \\ \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & \frac{2}{3} \end{bmatrix}.$$



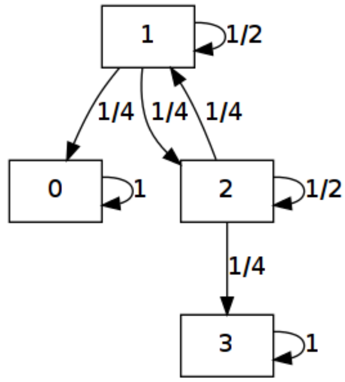
This is an irreducible chain, since every state can move to every other state. The unique steady-state distribution across the 3 states is $\vec{p}(\infty) = [2/7, 2/7, 3/7]$.

(b)

$$A := \begin{bmatrix} \frac{1}{2} & 0 & 0 & \frac{2}{3} \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{3} \end{bmatrix}$$

(c)

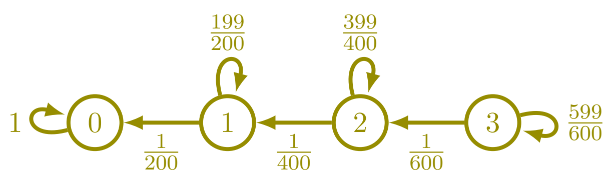
$$A := \begin{bmatrix} 1 & \frac{1}{4} & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{4} & 0 \\ 0 & \frac{1}{4} & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{4} & 1 \end{bmatrix}$$



This is a reducible chain, since once you reach the first or last state, you stay there forever. For this problem, there is no unique steady-state distribution. Instead, where you end-up depends in-part on where you start. $\vec{p}(\infty) = [c, 0, 0, 1-c]$ for some value of c that depends on the information about the initial state. We can see this from the directed graph of transitions, since if we end up in state 0 or state 3, there is no way to leave it ever again.

3. The breakdown of machines and structures are often modelled using Markov chains. One example of this is the deterioration of bridges. Imagine a new rope bridge in the Andes mountains held up by 3 jut ropes. Every time a person crosses the bridge, there is a chance that one of the ropes will break. Let a_n be the probability that one of the ropes breaks when there are n ropes, and assume that two ropes never break at once. For our purposes, $a_n = \dfrac{1}{200n}$. The bridge will collapse when there are no ropes left holding it up.

   (a) Draw graph of this Markov chain. Label all nodes with their corresponding state, and all edges with their transition probability. Discuss any uncertainties in the model structure and parameterization.



   (b) Construct a transition matrix for changes in the bridge's state every time a person crosses the bridge.

$$\begin{bmatrix} \frac{599}{600} & 0 & 0 & 0 \\ \frac{1}{600} & \frac{399}{400} & 0 & 0 \\ 0 & \frac{1}{400} & \frac{199}{200} & 0 \\ 0 & 0 & \frac{1}{200} & 1 \end{bmatrix}$$

   (c) If the Markov chain begins when the bridge is first built, what is the initial condition?

$$\vec{p}(0) = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

   (d) What is the probability that the bridge will have collapsed by the time 1000 people have crossed it? (Hint: Use python to perform this calculation. The function `linalg.matrix_power` will be useful. Be sure to explain your answer.)

The probability of bridge failure is $[1, 0, 0, 0] \cdot A^{1000}\vec{p}(0) = 0.47522759$.

4. In a manufacturing setting, Markov chains can be useful for describing throughput for certain processes. One example of this is a repair process at an electronics manufacturer. For this process, that states of the chain are called "stages". A device enters through "receive" stage and then is moved to the "debug" stage. Depending on the test results from "debug", it is moves to "scrap" or "repair" stages. From "repair", it moves to "test". From the "test" stage, the device will move to either "ship" or back to "debug". About 40% of the time, the debugging stage fails and the device is scrapped. About 80% of the time, the testing phase succeeds and the device is shipped. The average processing times for each transient stage are 10 minutes in receiving, 1 hour in debugging, 30 minutes in repair, and 15 minutes in testing. (The "scrap" and "ship" stages are absorbing stages, and formally have infinite processing times.)

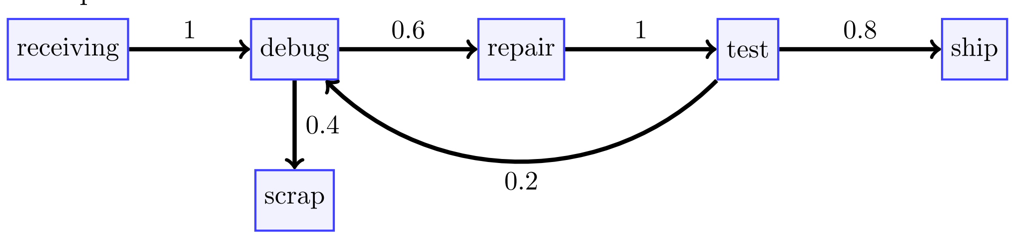
   (a) What is the probability that a device in the debug stage eventually returns to the debug stage?

   Modified from (Jonas, 2014) The probability of following this cycle path is $r = 0.6 \times 1 \times 0.2 = 0.12$.

   (b) What is the expected number of times a device will enter the debug stage? (Hint: the number of returns is geometrically distributed)

   If n is the number of returns to the debugging stage, geometric distribution gives $P(n) = r^n(1-r)$ where $r = 0.12$, as calculated above. The expected value of a geometric distribution is $r/(1-r) = 0.136$. But we didn't count the first visit – everything enters the debug stage at least once. So, a device enters debugging 1.136 times on average.

   (c) The total load of a device is the expected number of hours needed to handle each device before it is scrapped or shipped. Calculate the total load under the above model description.



   The most direct way to answer this is using the previous calculation. A device spends 10 minutes in receiving, and then does the 105 minute debug cycle 1.136 times, so $10 + 1.136 \times 105 \approx 129$ minutes per device.

5. Suppose you want to continuously represent the position of a point on the surface of a cube. If you want, you can assume the cube's radius is 1. Find at least 3 different ways to represent this point.

Without lose of generality, let's take the cube to be centered at the origin. We could use 3 real numbers (x, y, z) to represent a direction vector in 3 dimensions, with the extra condition that $(\max\{x, y, z\})^2 = 1$ always. We could use spherical coordinates (r, ϕ, θ) with a normalized radius, or cylindrical coordinates with (r, θ, z), also normalized. Third, we could use an integer 1 to 6 to determine the face, and then two real numbers for the position on the face.

6. Suppose we need to represent the positions of three carts on a simple railroad track with a station at each end, and no switches in between.

   (a) If we use a model space that is the position of each cart allong the track, $(x, y, z) \in [0, L]^3$ where L is the distance allong the track from the initial station to the final station, what constraints will the states always satisfy?

We can use three real coordinates, one for each cart. But we have to make sure the order of the carts never changes, probably by inforcing a minimum distance is always maintained, so that $0 \leq x < y < z \leq L$.

(b) If, instead, we represent the positions of the carts such that u is the distance from the initial station to the first cart, v is the distance from the first cart to the second cart, and w is the distance from the second to the third, what alternative constraints must (u, v, w) satisfy?

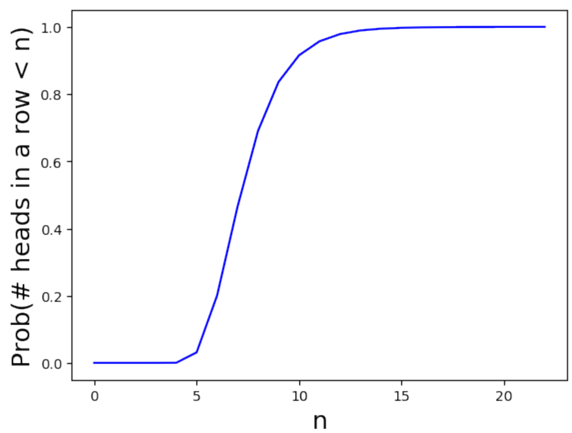
When we use relative positions u,v, and w, the no-passing condition means $0 < v$ and $0 < w$, while the range contraints implies $0 \leq u$ and $u + v + w \leq L$.

(c) Find conversion formulas to convert one state space to the other and vice-versa.

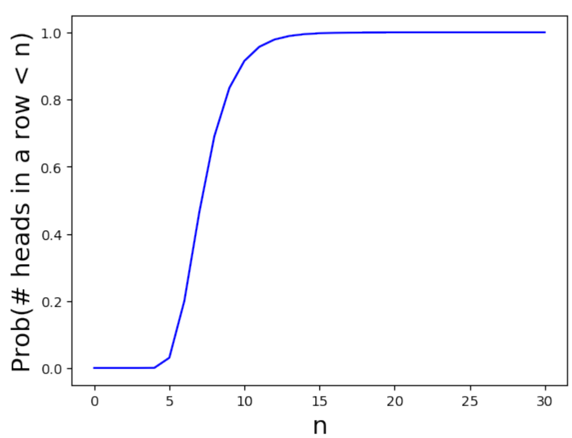
From (a) to (b), $u, v, w = (x, y - x, z - y)$. From (b) to (a), $x, y, z = (u, u + v, u + v + w)$.

7. Using Monte Carlo simulation, estimate the distribution for the maximum number of sequential heads in 100 flips of a fair coin. Perform at least 100,000 simulation runs. Present your results as a plot the number of runs with fewer than n heads in a row as a function of n.

You should get something that roughly looks like this:

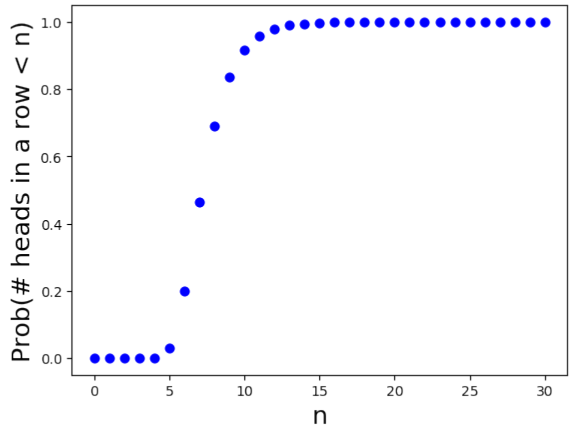


(that's fine for full marks).
Observe that it's still a bit choppy. You may think that you need even more simulations to smooth it out. For 10^7 simulations,



First notice that the x-axis is longer: with 10^7 simulations there are a few rare simulations with longer chains (1 with a chain of 30!). So this does capture rare events ($O(10^{-6})$) better.

But it's still choppy! That's because when we use a line, we think "continuous" but this is actually discrete. We don't "fill in" additional values between 4 and 5 – there are fewer than 4 heads or fewer than 5 heads. May be better for sanity to use symbols,



What do you think it would look like if you were looking for heads *or* tails? Test your hypothesis by trying it!

8. Imagine you start with an urn containing 2 stones, one black and one white. Reach into the urn, pick out a stone, then return the stone and a second stone of the same color as the one you picked out (so now there are 3 stones in the urn). Repeat this over and over until there are 1,000 stones in the urn.

   (a) What do you expect to have happened to the stones in the urn? Will most of the stones in the urn be the same color? Or will do you think it's more likely that the ratio of white to black stones will be 1 to 1? Or something different?

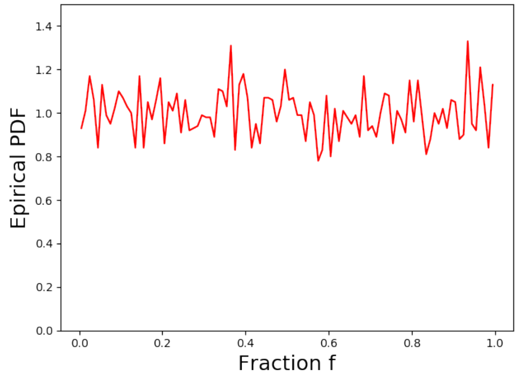
   As long as your hypothesis is reasonable, it's fine.
   I initially thought you'd get like for like – that is, if you first picked some black stones, then you'd preferentially pick black stones, and vice-versa. So most of the stones would end up being black or white, and the ratio would be some n:1 where $n \neq 1$.

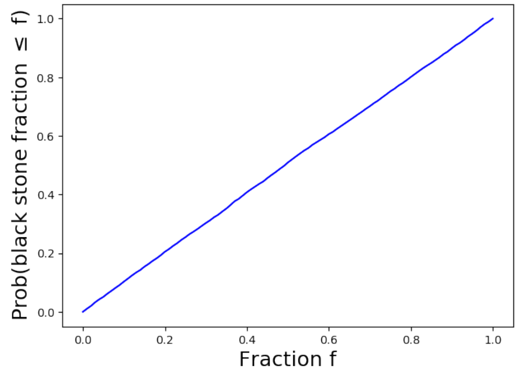
   (b) What should the empirical PDF and CDF of the outcome look like under your prediction for part a?

   As long as this matches (a), it's fine.
   My hypothesis would suggest a "bimodal" distribution in the fraction – either the fraction is low or high.

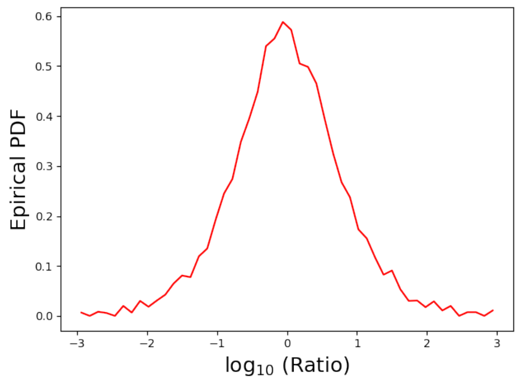
   (c) Simulate 10,000 runs of this process, and plot the empirical PDF and CDF.

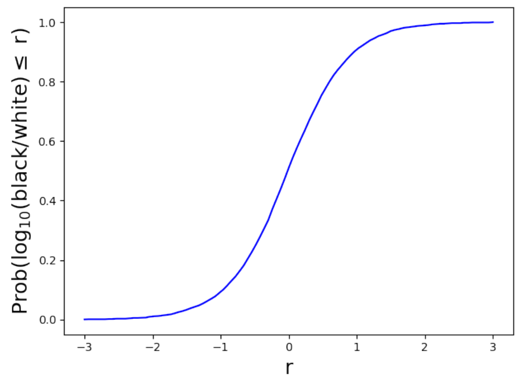
   Start with the fraction of one of the stone colors – say, the black stones (the fraction of the white stones is the fraction of the black stones subtracted from 1):





(using 101 bins to compute the empirical PDF; fine if used up to 1001 bins, i.e., 0 to 1000). Let's also consider the ratio of black stones to white stones – using the log to zoom in a bit:

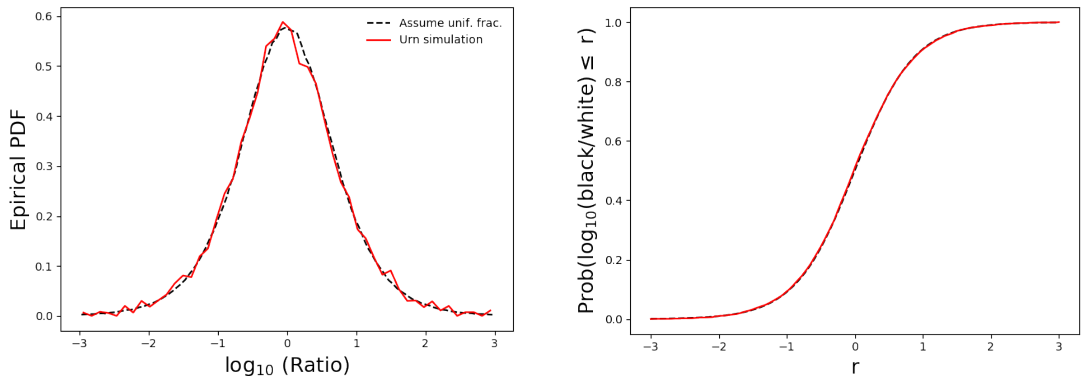




(using 51 bins to compute the empirical PDF). Note that the empirical PDF is noisier than the empirical CDF – this is why people often consider the CDF, by integrating the PDF you smooth out some of the noise.

(d) Discuss.

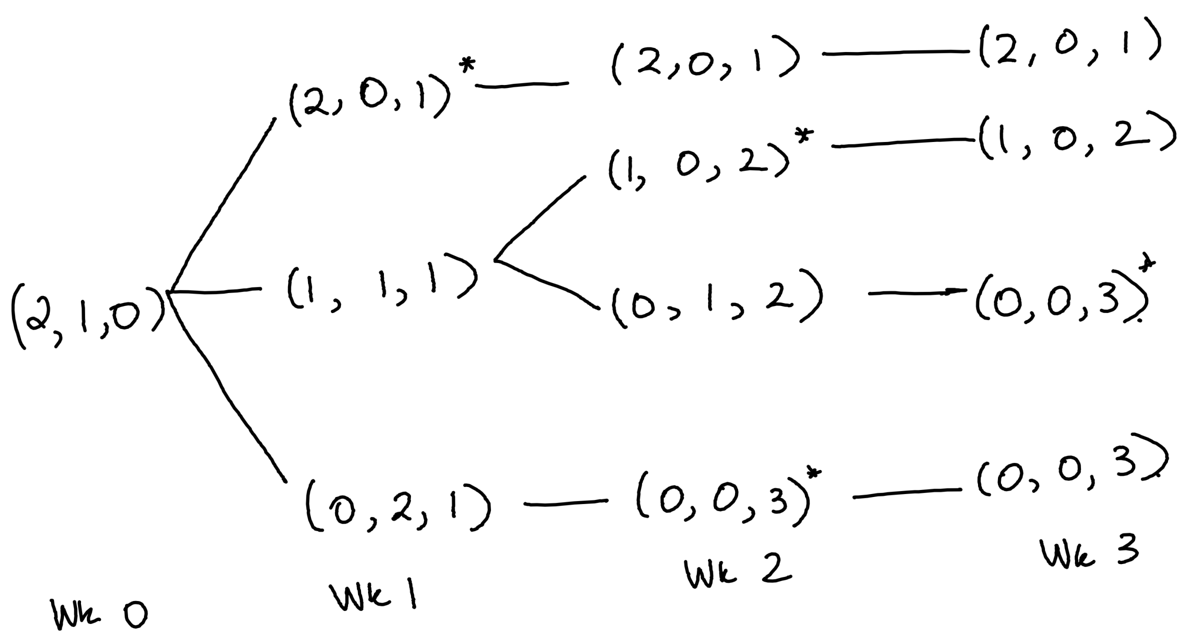
Well first off, from the PDF/CDF of the fraction of black stones, it's pretty clear that the distribution on fraction of black stones (also white stones) looks to be **uniform**! That is, with equal probability you can get any fraction (black)/(black+white) or (white)/(black+white) between 0 and 1.
The ratio black stones/white stones is more interesting. It looks like the median and mean of the log ratio is roughly 0, suggesting that the expected ratio is 1/1. But there is variability in the ratio! With some small probability you can indeed get mostly white or mostly black. Is this consistent with a true uniform distribution? Can we use the ratio to further verify that the distribution is uniform? Yes. We can sample $x =$(black stones)/(black stones + white stones)$\sim$ Unif[0,1]. Then the ratio of black stones/white stones is $x/1-x$.



Good support for the hypothesis that the fraction of black stones is uniformly distributed!

9. Greenwood and Yule proposed the following model of disease transmission in 1920. Disease transmission is broken into discrete generations, and the number of infectious people in each generation is a binomial random sample from the currently susceptible people. People can only get the disease once, so people sick this week are recovered and resistant against re-infection next week.

   Let's call the model space (S, I, R) tracking the number of susceptibles, infecteds, recovereds. Then the initial state is $(2, 1, 0)$. The following figure describes the progress over the weeks (The * indicate the point at which the outbreak has ended):



   (a) Suppose a family has 1 sick kid with measles and 2 susceptible kids. What are the 3 possible states next week and their probabilities?

   From the figure we see the possible states are $(2, 0, 1)$, $(1, 1, 1,$ $(0, 2, 1)$.
   Assume that the probability that an infected sibling transmits to a susceptible is p. Then

   - $(2, 0, 1)$ occurs with probability

$$\binom{2}{0} p^0 (1-p)^{2-0} = (1-p)^2$$

- $(1,1,1)$ occurs with probability

$$\binom{2}{1} p^1(1-p)^{2-1} = 2p(1-p)$$

- $(0,2,1)$ occurs with probability

$$\binom{2}{2} p^2(1-p)^{2-2} = p^2$$

As a sanity check, observe that our probabilities sum up to 1,

$$p^2 + 2p(1-p) + (1-p)^2 = 1.$$

as they should.

(b) What are the family's 4 possible states the second week, and their probabilities?

From the figure we see the possible states are $(2,0,1)$, $(1,0,2)$, $(0,1,2)$, and $(0,0,3)$. Assume that the probability that an infected sibling transmits to a susceptible is p. Then

- For $(2,0,1)$: $(2,0,1)$ occurs with probability $(1-p)^2$ as in the previous week – no change since there's no one left to infect!
- For $(1,0,2)$: $(1,1,1)$ occurs with probability $2p(1-p)$. Then going from $(1,1,1)$ to $(1,0,2)$ occurs with probability

$$\binom{1}{0} p^0(1-p)^{1-0} = (1-p).$$

  Therefore the probability of $(1,0,2)$ is $2p(1-p) \times (1-p) = 2p(1-p)^2$.
- For $(0,1,2)$: $(1,1,1)$ occurs with probability $2p(1-p)$. Then going from $(1,1,1)$ to $(0,1,2)$ occurs with probability

$$\binom{1}{1} p^1(1-p)^{1-1} = p.$$

  Therefore the probability of $(0,1,2)$ is $2p(1-p) \times p = 2p^2(1-p)$.
- For $(0,0,3)$: $(0,2,1)$ occurs with probability $= p^2$. But there's no one left to infect – the next week HAS to be $(0,0,3)$. Therefore the probability of $(0,0,3)$ at week 2 is p^2.

As a sanity check, observe that our probabilities sum up to 1,

$$(1-p)^2 + 2p(1-p)^2 + 2p^2(1-p) + p^2 = 1,$$

as they should.

(c) What are the probabilities that when the outbreak ends, the family had a total of 1,2, or 3 cases of measles?

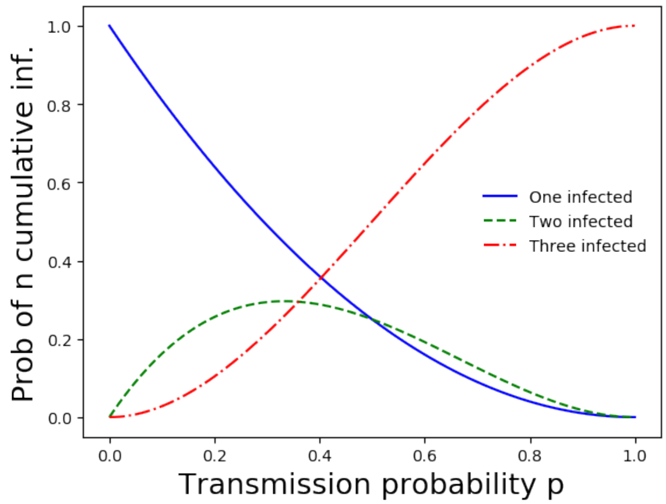
- Probability of 1 infection: $(2,0,1)$ occurs with probability $(1-p)^2$ (computed in (a)).
- Probability of 2 infections: $(1,0,2)$ occurs with probability $2p(1-p)^2$ (computed in (b)).
- Probability of 3 infections: $(0,0,3)$ occurs with probability p^2 (achieved in week 2, see (b)) + $2p^2(1-p)$, which is the probability of being at $(0,1,2)$ in week 2, that last person has to recover (computed in (b)). Thus $(0,0,3)$ occurs with probability $p^2(3-2p)$.

As a sanity check, observe that our probabilities sum up to 1,

$$p^2 + 2p(1-p)^2 + p^2(3-2p) = 1.$$

as they should.

So what does this mean? Can graph it:



For low-probability transmission (p near 0), it's most likely there will be only one infection, maybe 2. For high-probability transmission (p near 1), it's most likely that all theh kids will get sick. But we also note that the structure, i.e. the chains creating the infection, create an asymmetry in the outcomes. For example, if $p = 0.5$, then the probability of 1 or 2 infections is 1/4, while the probablity or 3 infections is 1/2.