Statistical Analysis of Snakes and Ladders

Year 3 Numerical Modelling of Physical Systems

By Mia Boulter

1 Introduction

The aim of this project is to find the average number of turns to complete a game of snakes and ladders given any board, for each start position. Also included should be the variance on these values. Programmatically one should be able to generate a random board and obtain these values for that board. To achieve this the game will be modelled as a Markov chain, which describes a sequence of possible events where the probability of an event is dependent on only the state that followed the previous event. Following modelling the game, investigations will be done into how different finish conditions affect the players in various ways.

2 Theory

2.0 Establishing Rules

In snakes and ladders, a player has a chance of going from one square to another, with the snakes or ladders impacting these probabilities by shifting the probability of landing on one square on to a different square, but ultimately this doesn't impact the method used to calculate the aim of the experiment.

The game can have various rules including:

1. Start **on** first square or **off** the board.
2. Get a bonus dice roll should you get a max roll.
3. Need to roll the exact distance to the finish in order to finish.
4. If you roll a greater number than the distance to the finish, move back a number of squares equal to the difference in the distance to the finish and your roll.

For this explanation, we will use the rules:

1. Start **off** the board.
2. No bonus dice roll.
3. Roll the exact distance to finish
4. No moving back if you roll more than the distance to the finish.

The explanation would apply to any other board and rule set except if you were to use the bonus roll rule since the possible places to reach in one turn makes generating a transition matrix much more complex.

2.1 The Transition Matrix

For Markov chains like this, one uses a transition matrix (sometimes referred to as a transfer matrix or stochastic matrix) to describe the transitions between states. For a snakes and ladders board like this (where the zeroth square implies starting off the board, and the ninth square is the finishing square):

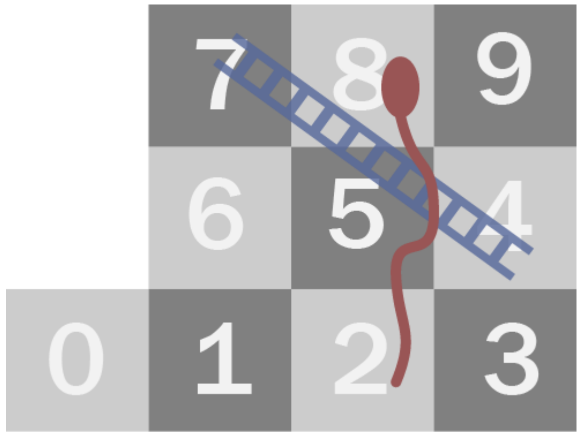

Figure 1 – Example Snakes and Ladders Board

The transition matrix would contain the probabilities of moving from each square to another, with rows being the square the player is moving from, and columns being the square the player is moving to. For the board in figure 1, the transition matrix is as follows.

$$\boldsymbol{P} = \begin{bmatrix} 0 & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & 0 & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & 0 & 0 \\ 0 & 0 & \frac{1}{6} & \frac{1}{6} & 0 & \frac{1}{6} & \frac{1}{6} & \frac{1}{3} & 0 & 0 \\ 0 & 0 & \frac{1}{6} & \frac{1}{6} & 0 & \frac{1}{6} & \frac{1}{6} & \frac{1}{3} & 0 & 0 \\ 0 & 0 & \frac{1}{6} & 0 & 0 & \frac{1}{6} & \frac{1}{6} & \frac{1}{3} & 0 & \frac{1}{6} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{6} & 0 & 0 & \frac{1}{3} & \frac{1}{6} & \frac{1}{6} & 0 & \frac{1}{6} \\ 0 & 0 & \frac{1}{6} & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{6} & 0 & \frac{1}{6} \\ 0 & 0 & \frac{1}{6} & 0 & 0 & 0 & 0 & \frac{2}{3} & 0 & \frac{1}{6} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

This would be calculated simply by placing the player on each valid square (i.e. not where a ladder start, or a snake head are.) and trying each possible result of a dice roll to find the chance of going to each square. By definition, each row corresponding to a state that it is possible to be in must sum to one. For states not possible to be in, all entries would be zero. These could be removed, but it's not necessary to do so, they are ignorable.

A snakes and ladders board is an example of an absorbing Markov chain, that is, the transition matrix contains both transient and absorbing states. Once an absorbing state is entered, it cannot be left; as opposed to transient states which can be left.

The transition matrix for an absorbing Markov chain can be brought into canonical form, with the transient states first, then the absorbing states last,

$$\mathbf{P} = \begin{array}{c c} & \begin{array}{cc} \text{TR.} & \text{ABS.} \end{array} \\ \begin{array}{c} \text{TR.} \\ \text{ABS.} \end{array} & \left(\begin{array}{c|c} \mathbf{Q} & \mathbf{R} \\ \hline \mathbf{0} & \mathbf{I} \end{array}\right) \end{array}$$

This can be done by checking if the state contains the number 1. This would imply it's an absorbing state because there is a probability of 1 of it staying in that state. Then you generate a new arrangement from this, then cast that arrangement onto each column and row of the matrix. One can do this by passing a list of the new arrangement into the __getitem__ method of a NumPy array of each row, then each column.

Since the absorbing state for the snakes and ladders board is the final square, the transition matrix is already in this form. From here we can extract Q, which will hold probabilities of transition for only transient states. For our example,

$$\boldsymbol{Q} = \begin{bmatrix} 0 & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & 0 & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & 0 \\ 0 & 0 & \frac{1}{6} & \frac{1}{6} & 0 & \frac{1}{6} & \frac{1}{6} & \frac{1}{3} & 0 \\ 0 & 0 & \frac{1}{6} & \frac{1}{6} & 0 & \frac{1}{6} & \frac{1}{6} & \frac{1}{3} & 0 \\ 0 & 0 & \frac{1}{6} & 0 & 0 & \frac{1}{6} & \frac{1}{6} & \frac{1}{3} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{6} & 0 & 0 & \frac{1}{3} & \frac{1}{6} & \frac{1}{6} & 0 \\ 0 & 0 & \frac{1}{6} & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{6} & 0 \\ 0 & 0 & \frac{1}{6} & 0 & 0 & 0 & 0 & \frac{2}{3} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

2.2 Finding the Average

To find the probability of being in each of the transient states j after n turns, given an initial state i, the row in the matrix, we do $\boldsymbol{Q}^n$.

The average number of turns spent on each state j after n turns (for each initial state i) will be:

$$\sum_{m=0}^{n} \boldsymbol{Q}^m$$

An absorbing state will inevitably be reached,

$$\boldsymbol{Q}^n \to \boldsymbol{0} \text{ } as \text{ } n \to \infty$$

Therefore, the sum converges with $n \to \infty$. From this an equation for $\boldsymbol{N}$, the value which the sum tends towards, can be developed

$$\boldsymbol{N} = \sum_{m=0}^{\infty} \boldsymbol{Q}^m = \boldsymbol{I} + \boldsymbol{Q}\left(\sum_{m=0}^{\infty} \boldsymbol{Q}^m\right) = \boldsymbol{I} + \boldsymbol{QN}$$

And so, one obtains the equation

$$\boldsymbol{N} = (\boldsymbol{I} - \boldsymbol{Q})^{-1}$$

$\boldsymbol{N}$ is known as the fundamental matrix. As shown, it is given by the inverse of $\boldsymbol{Q}$ subtracted from the identity matrix. An item at position (i, j) in N gives the average number of times in state j given an initial state i, before an absorbing state is reached. For our example,

$$\boldsymbol{N} = \begin{bmatrix} 1.000 & 0.167 & 1.433 & 0.433 & 0.000 & 0.758 & 1.264 & 3.544 & 0.000 \\ 0.000 & 1.000 & 1.400 & 0.400 & 0.000 & 0.700 & 1.167 & 3.733 & 0.000 \\ 0.000 & 0.000 & 2.400 & 0.400 & 0.000 & 0.700 & 1.167 & 3.733 & 0.000 \\ 0.000 & 0.000 & 1.200 & 1.200 & 0.000 & 0.600 & 1.000 & 3.200 & 0.000 \\ 0.000 & 0.000 & 0.000 & 0.000 & 1.000 & 0.000 & 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 1.200 & 0.200 & 0.000 & 1.850 & 1.083 & 2.867 & 0.000 \\ 0.000 & 0.000 & 1.200 & 0.200 & 0.000 & 0.350 & 2.583 & 2.867 & 0.000 \\ 0.000 & 0.000 & 1.200 & 0.200 & 0.000 & 0.350 & 0.583 & 4.867 & 0.000 \\ 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 1.000 \end{bmatrix}$$

It is easy to see that column 0 is as expected. For our board, if you start on square 0, you have been in that state once, then after that there is no way to return, so the average number of turns spent on that square is 1. If you start on a different square, there is no way to get to square 0, so the average number of turns spent there will be zero.

Finally, to get the average number of turns to reach an absorbing state, we multiply by a column vector of which all the items are 1.

$$\boldsymbol{c} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

This effectively gives a vector of the same shape as $\boldsymbol{c}$, where each item is the sum of the corresponding row in $\boldsymbol{N}$. By adding the expected times in each state for a given initial state, we obtain the average number of turns to reach an absorbing state. For our example this is

$$\boldsymbol{t} = \begin{bmatrix} 8.6 \\ 8.4 \\ 8.4 \\ 7.2 \\ 1.0 \\ 7.2 \\ 7.2 \\ 7.2 \\ 1.0 \end{bmatrix}$$

So, for a start position 0 as in the board, 8.6 is the average turns that a game will take.

With a transition matrix where it is not already in canonical form it is important to keep track of what state each of the rearranged rows corresponds to.

From $\boldsymbol{t}$ one can now easily order the squares by the average number of turns until victory (4 and 8 are ignored as it is not possible to be in those states).

Avg. Turns until victory	**Current Square(s)**
7.2	7, 6, 5, 3
8.4	2, 1
8.6	0

2.3 Variance

The following is a graph plotted of the distribution of number of turns taken to finish from starting on square 0 following running 1,00,000 games on the example board. This yielded a numeric average turn number of 8.604454 with the variance on this value being 48.067195.

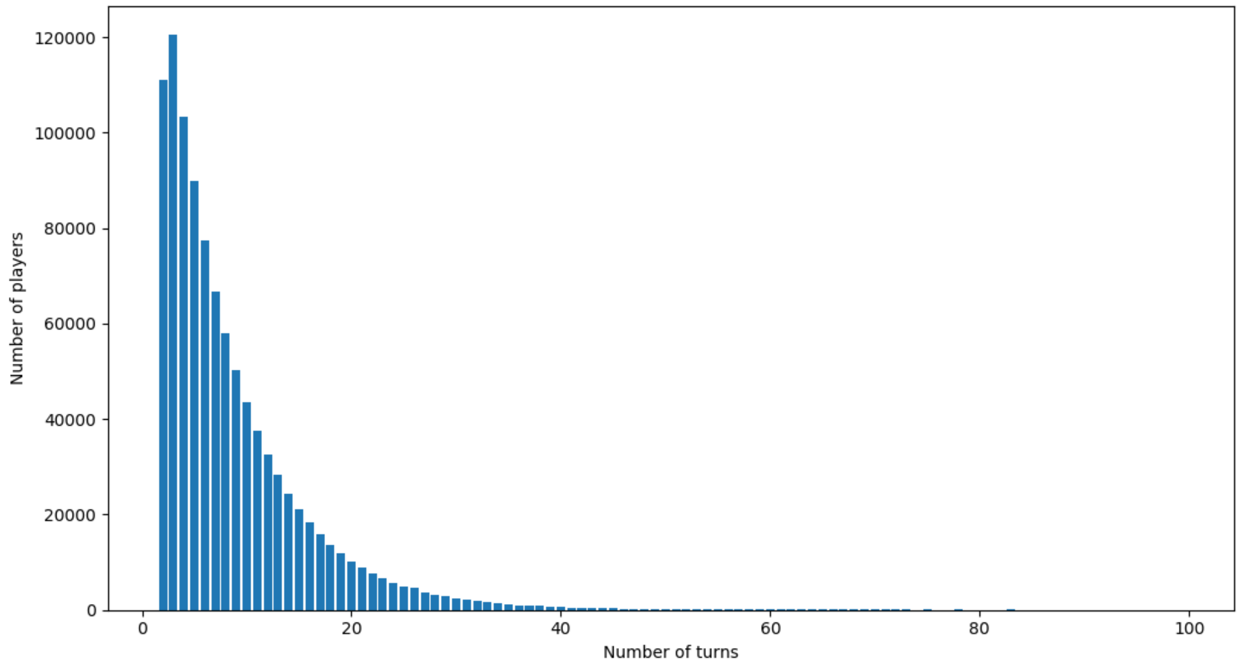

Figure 2 – Plot of number players who take each number of turns to complete a game from a sample of 1,000,000 players. A trend is visible but unclear on the lower end because of the number of turns being integer values which has a significant impact for this small board size.

Analytically, the variance on the number of turns to finish for each start position is given by

$$\boldsymbol{\sigma}^2 = (2\boldsymbol{N} - \boldsymbol{I})\boldsymbol{t} - \boldsymbol{t} \circ \boldsymbol{t}$$

$\boldsymbol{t} \circ \boldsymbol{t}$ is the Hadamard product of $\boldsymbol{t}$ with itself, that is, the item (i, j) in the resultant matrix is the items (i, j) in the two matrices multiplied together.

For our example this is

$$\boldsymbol{\sigma}^2 = \begin{bmatrix} 47.92 \\ 47.76 \\ 47.76 \\ 47.52 \\ 00.00 \\ 47.52 \\ 47.52 \\ 47.52 \\ 00.00 \end{bmatrix}$$

3 Further Analysis at Larger Scale

So far only a small board has been analysed, but this has its limits in how well we can visualise distributions. This is noticeable in figure 2, where I noted that trend on the lower end is unclear. From here on we will use a different, much larger board.

The board would be generated by the code using the following

```
Board({'squares': 100, 'snakes': {97: 55, 68: 3, 5: 2, 25: 7, 78: 37,
79: 1, 14: 6, 45: 10, 39: 21}, 'ladders': {88: 93, 15: 70, 74: 95, 67:
71, 84: 86, 24: 46, 28: 31, 34: 76}, 'start_off_board': True})
```

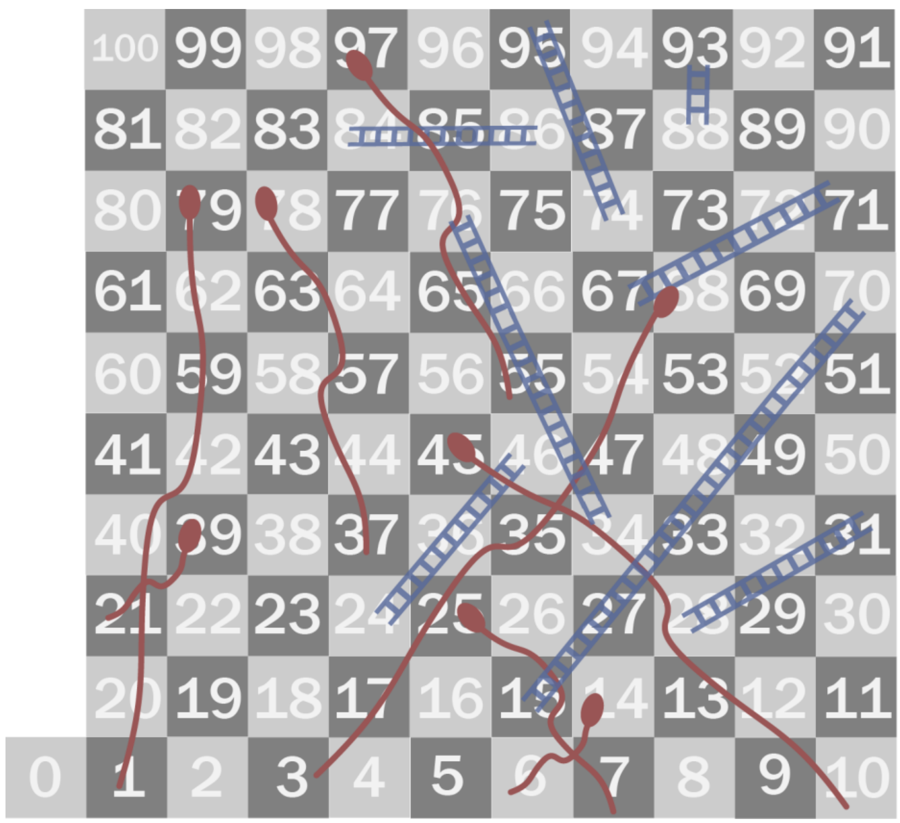

Figure 3 – A larger snakes and ladders board

3.0 Uses of the Transition Matrix

The transition matrix can be used to find the probability of completion after a certain number of turns. If you take Q, the transient portion of the transition matrix, to the power of the number of turns, you have the probability of being on each transient square. Then you can take one minus the sum of these values for a given start square to find the cumulative probability of absorption.

Taking 0 as the start square, I took a version of figure 2 for this board, for 100,000 players. On there also I plotted the differential of the cumulative probability found to have finished in order to find the number of players finished on that turn specifically. Scaling the prediction to the number of players, the following plot shows how the statistical trial lines up with the analytical prediction.

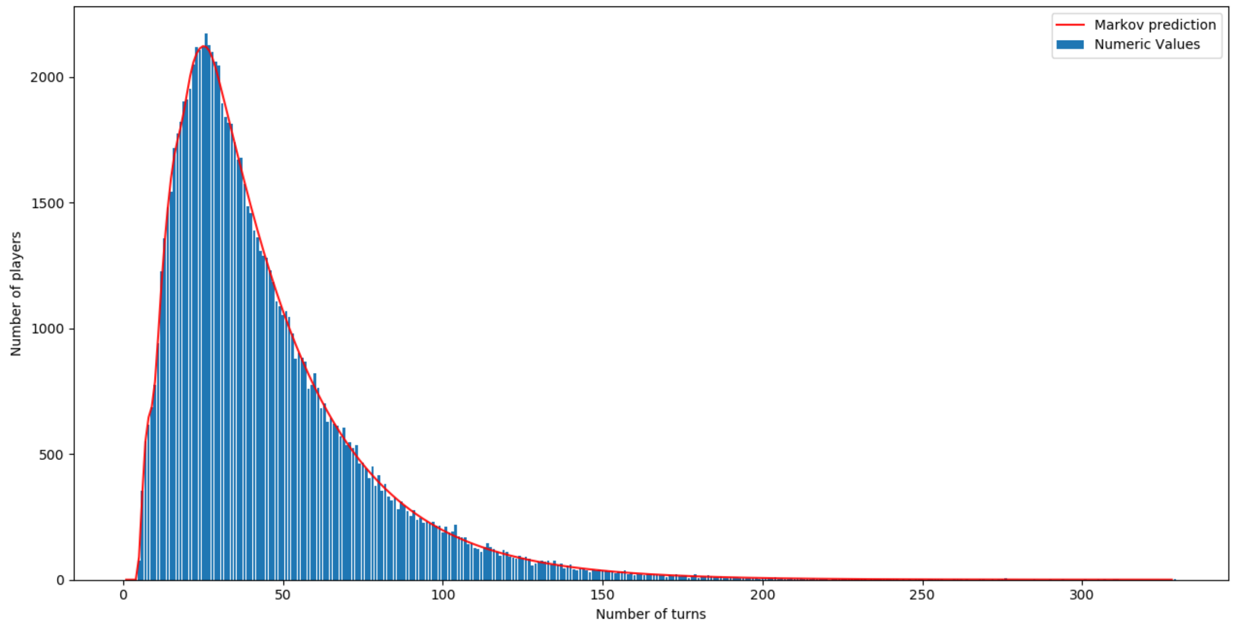

Figure 4 – Plot of number players who take each number of turns to complete a game from a sample of 100,000 players. An analytical prediction from the transition matrix is shown in red.

To assess the disorder of the system at various points in time, one can find the entropy. The way the entropy changes with time can be observed. Unlike a thermodynamic system, in the case of a system containing absorbing states, entropy can decrease – the players are locked into the absorbing states and so over time order returns.

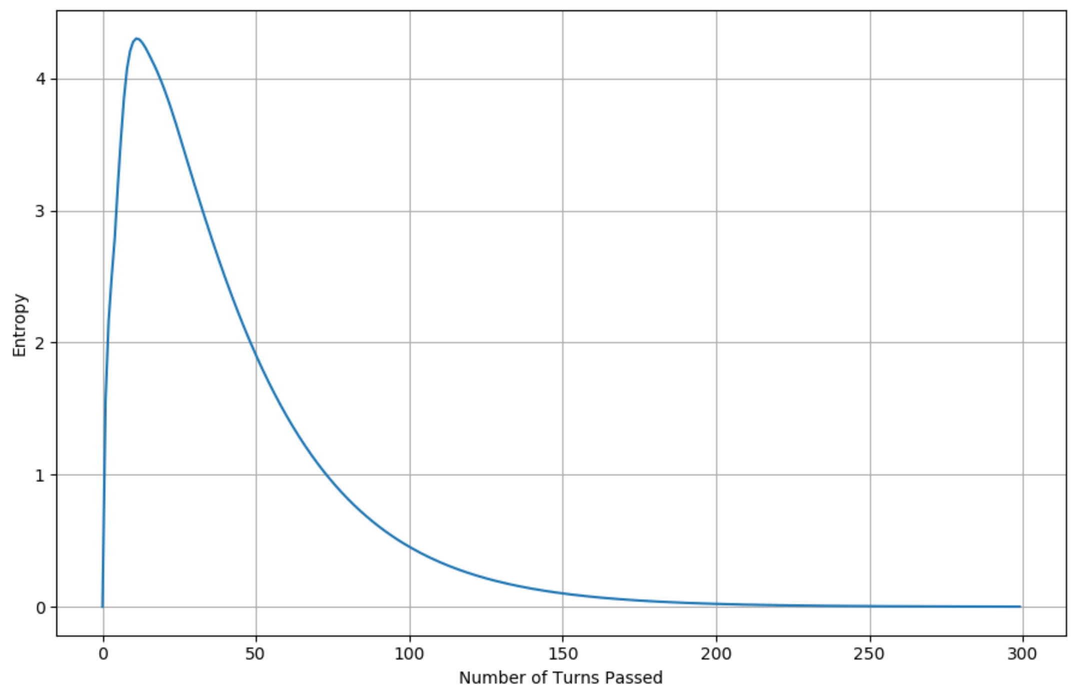

Figure 5 – Plot of how entropy varies with time, with entropy obtained by assessing the transition matrix at each turn.

The entropy shows a similar shape to the probability of having completed a game. Though it peaks earlier, at turn 11. The players start at the same place, then spread out, causing the initial increase in entropy. The number of players finishing the game starts to become significant after turn 11. For snakes and ladders, entropy settles at zero because all players are reach one single absorbing state.

We can consider how the average number of turns to finish varies by square for this board.

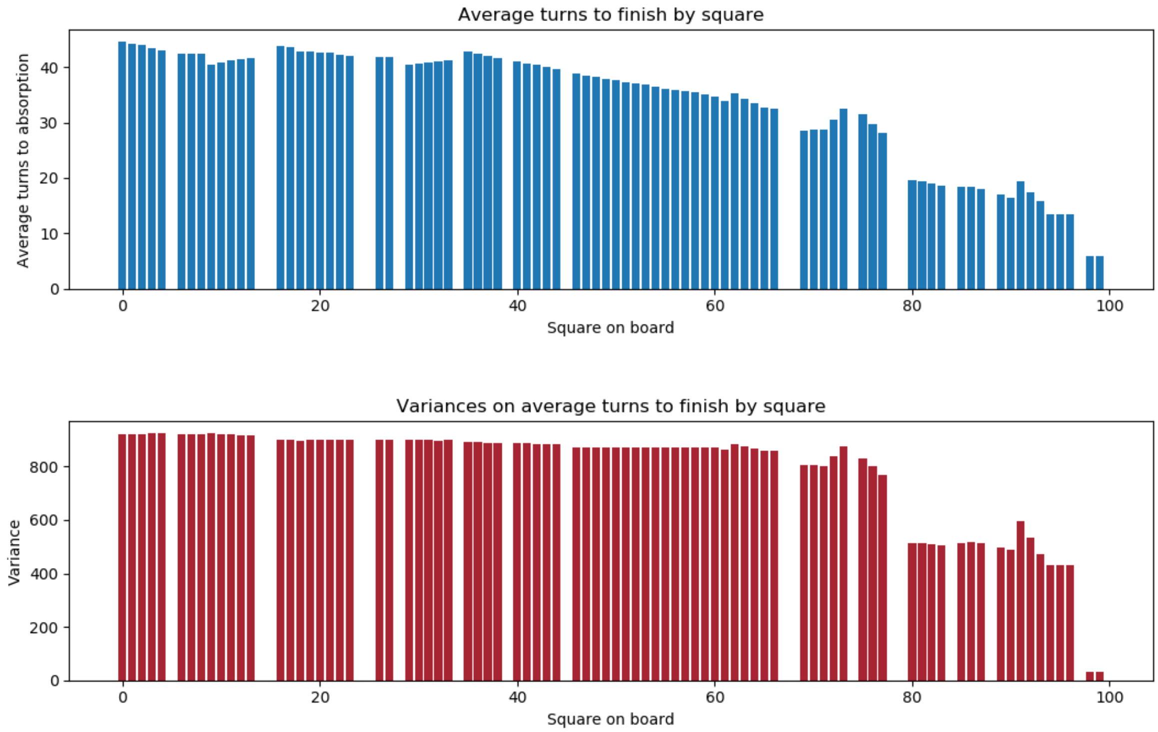

Figure 6 – Plot of average number of turns to absorption by square, along with the variances.

The board behaves in a very complex way making it difficult to comment on this graph. I will comment on one part with a possible explanation. The average turns to finish on square 62 increases from square 61, this may be because moving from 61 to 62 puts a snake (square 68) in range if the player rolls a 6; while being on 61, they have an extra chance to hit the ladder at square 67, bypassing the snake. On square 63 the average number of turns is lower again because a chance to roll a 6 and bypass the snake is introduced. It is interesting that the variances seem to change in steps, rather than progressively like the average turns to finish do, and that pattern seems to be broken on squares greater than 60.

3.1 The Effects of Different Finish Conditions

Next, I will investigate how finish conditions affect the game. The three possible finish conditions are as follows:

- Exact finish – The player must land on the final square exactly, otherwise they stay on the square they are currently on.
- Reverse Overshoot – The player must land on the final square exactly. If they roll a number greater than what they need, they move to the final square, then backwards for the remaining amount of squares their roll determined they do.
- No Exact finish – If the player rolls a number greater than the distance to the final square, they move to the final square.

To start with I found the average turns to finish for a start position of zero.

```
-------------------------------------
Exact Finish
-------------------------------------
    Answer: 44.54342601137766
  Variance: 922.299209057067
Time taken: 0.01253470000000001 seconds
-------------------------------------

-------------------------------------
Reverse Overshoot
-------------------------------------
    Answer: 61.662488046261736
  Variance: 2460.428166002517
Time taken: 0.011930100000000055 seconds
-------------------------------------

-------------------------------------
No Exact Finish
-------------------------------------
    Answer: 38.705963135922886
  Variance: 789.3072518463007
Time taken: 0.012158499999999961 seconds
-------------------------------------
```

Immediately one can see the significant differences in variance. The reverse overshoot finish condition must have a big impact on the disorder. Looking at the entropy would be a good idea.

But before looking at the entropy I decided to how individual squares are affected using the fundamental matrix. For a start position of zero, we take the first row of the fundamental matrix. As explained previously, this contains the expected number of times each state is occupied before absorption. This reveals some interesting things with regards to the finish condition.

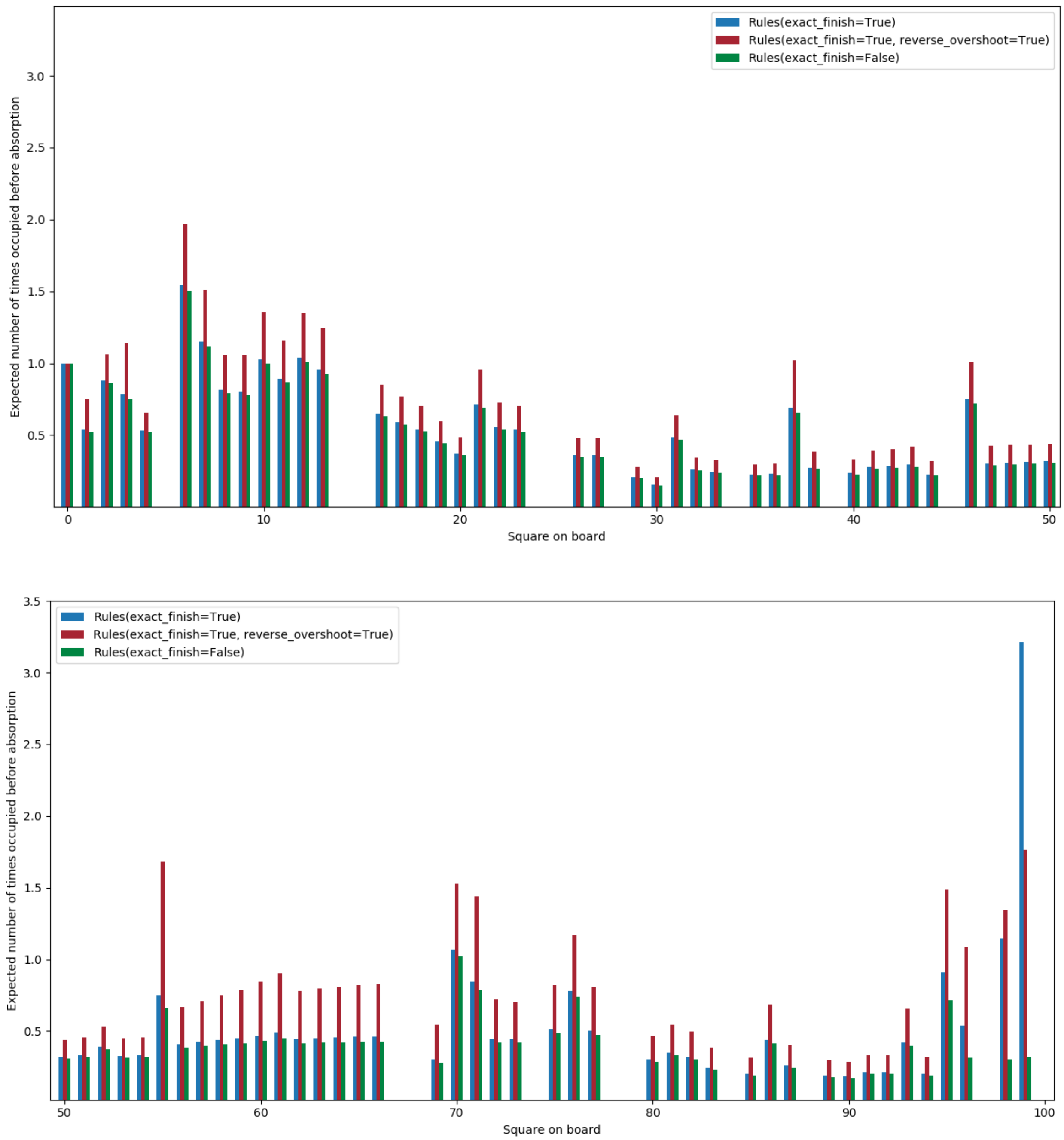

Figure 7 – Comparison of how three finish conditions affect the average number of times a state is occupied

In figure 7 we can see that each finish condition has an expected number of turns to be on square 0 of 1. This is expected since square zero is defined as being off the board, and so there is no way to return there after leaving. After that the finish conditions begin to result in some differences.

Exact finish is consistently higher than non-exact finish. This would be the result of the finish condition making you much more likely to hit the snake at square 97, which takes you back to square 55 where many snakes lie in your path to finishing. To finish you need to roll the exact number to reach the end, and if you are below square 97, you have just as much chance of hitting the snake as you do to finish. For this reason, you will likely visit most squares much more often.

For the most part, squares where a snake ends, or a ladder ends, have a higher expected number of visits. This makes sense because there is actually two squares on the board on which a player can land and end up on these squares.

The expected number of turns on square 99 is low for non-exact finish because you will be absorbed next turn if you land on that square. It's higher for exact finish than it is for reverse overshoot because with exact finish, a player on square 99 will stay on it until they finish, but reverse overshoot introduces a high chance to move off the square then finish without returning to that square.

Something else that can be investigated is the how the cumulative probability of absorption changes with turn number for a start position of zero. This is the integral of what we plotted as the prediction in figure prediction, that is, $1 - sum(\boldsymbol{Q}^n[0])$.

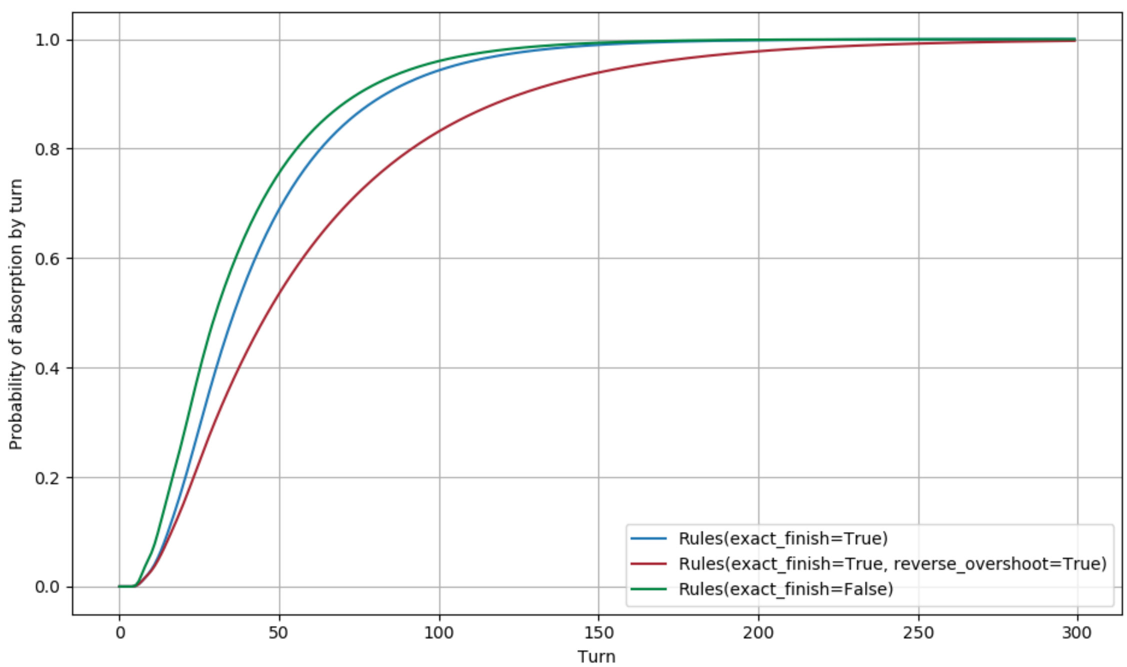

Figure 8 – Cumulative probability of absorption by turn number for three different finish conditions

Non-exact finish increases most steeply, exact finish is next in it's approach to 1 because the finish condition adds a chance to hit the snake on square 97, but reverse overshoot approaches 1 much slower because of the huge probability to hit the snake at 97.

Finally, we will look at how the finish condition affects the entropy.

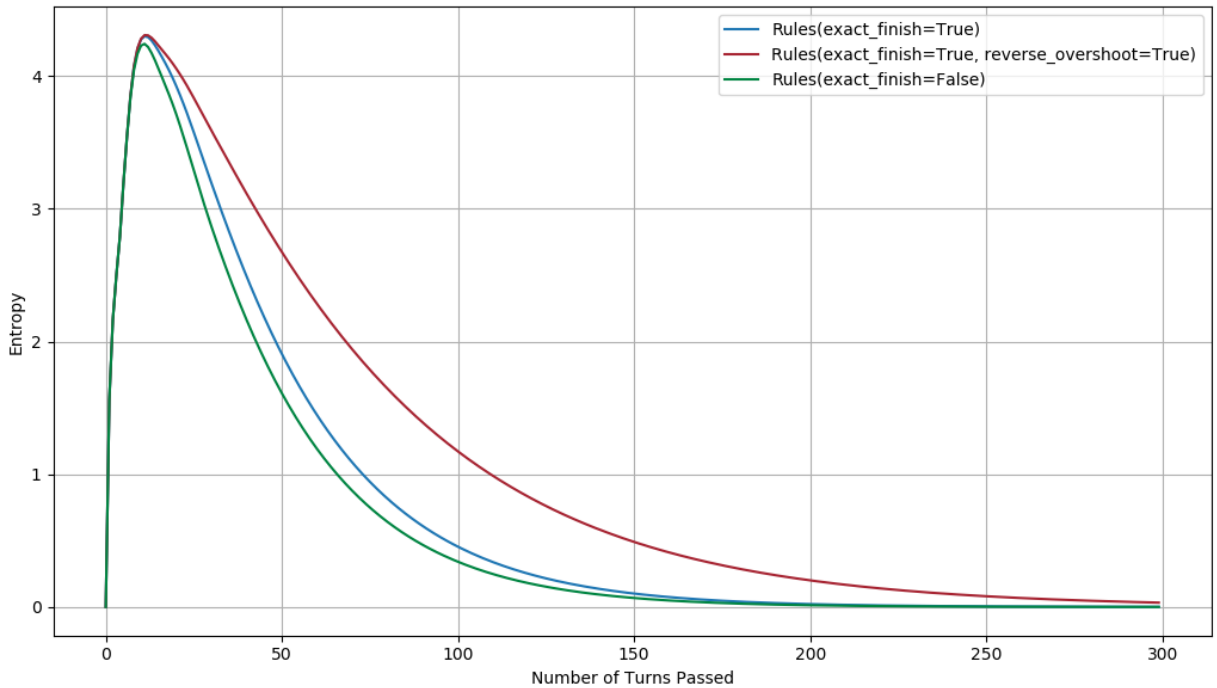

Figure 9 – The entropy to turn relation for different finish conditions.

The peak in entropy is reached at the same turn number for all finish conditions – turn 11. The peak in entropy is roughly the same value for exact finish and reverse overshoot, a value higher than that for non-exact finish. After that the entropy for exact finish decreases similarly to how non-exact finish does, while reverse overshoot decreases much more slowly. It makes sense that reverse overshoot causes these higher entropies since it randomly splits the players up even when they are near the finish.

As I final thing I was interested on how significant an impact the snake on square 97 has on the cumulative probability of finishing, and the entropy, for the various finish conditions. I removed the snake on square 97 and redid these plots.

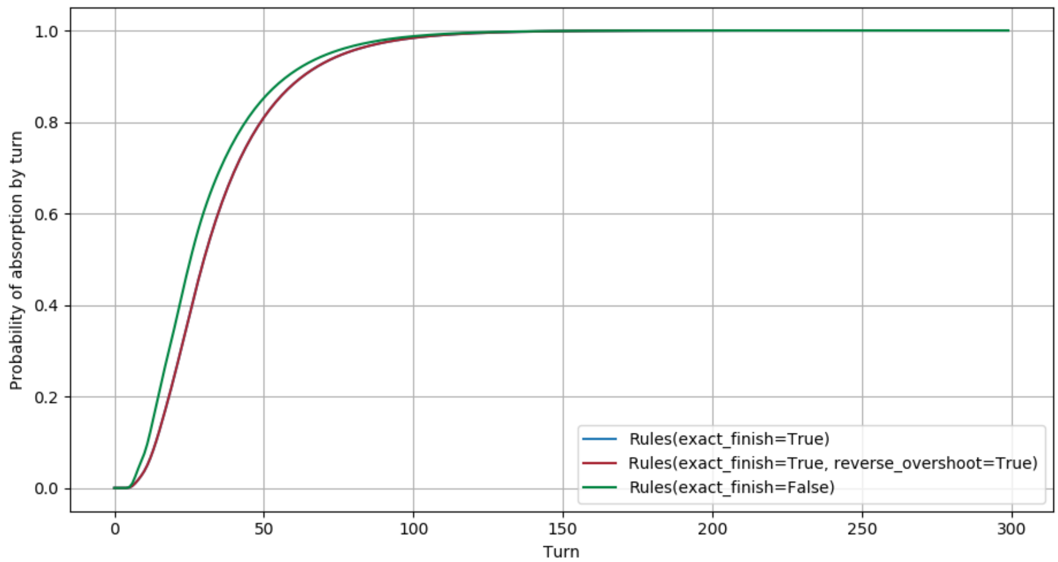

Figure 10 – How cumulative probability of absorption varies with snake at 97 removed

The cumulative probability for reverse overshoot is now in line with the exact finish. This is because, without the snake on square 97, they behave the same in the area close to the finish. Even if reverse overshoot applies, a player will still be left in a state in range of the finish, and so there is still a 1/6 chance to finish as there is in exact finish.

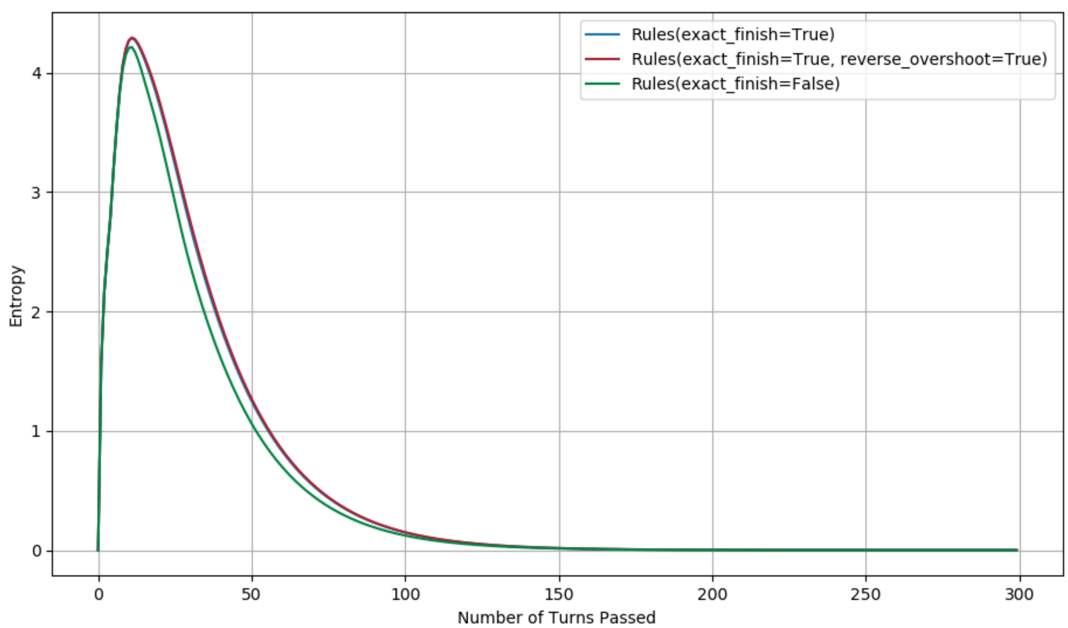

Figure 11 – How entropy varies by turn with snake at 97 removed.

The entropy of the two line up also since reverse overshoot can no longer maintain a high disorder like it could when the snake was present.

Conclusion

The process to find the average number of turns to finish a game of snakes and ladders begins with finding the transition matrix of the board. From this one extracts the transitory states and uses that reduced matrix to obtain the fundamental matrix. The fundamental matrix contains all the information needed to calculate the expected number of turns from each start state, and the variance on those values.

From there, for a large board, I analysed various things such as the average number of turns to finish, and how the entropy changes by turn. A notable discovery I made was the significant impact that snakes near the final square have on the reverse overshoot finish condition.

Bibliography

Grinstead, CM & Snell, JL, 1997, *Introduction to Probability: Second Revised Edition*, American Mathematical Society

Tolver, A, 2016, *An Introduction to Markov Chains*, University of Copenhagen