Markov Chain Simulations

Moti Ben-Ari

http://www.weizmann.ac.il/sci-tea/benari/

April 30, 2023

Contents

1 Introduction

Simulations are an excellent way of understanding probability, especially the behavior of processes of long duration. Simulation programs enable the user to perform experiments by varying the parameters of problems and analyzing. The simulations in this archive are of processes known as *Markov chains*, where the next state of the system depends only on the current state and not on the history of how the process got to the current state.

A knowledge of probability is assumed at the level of the first few chapters of [3, 8].

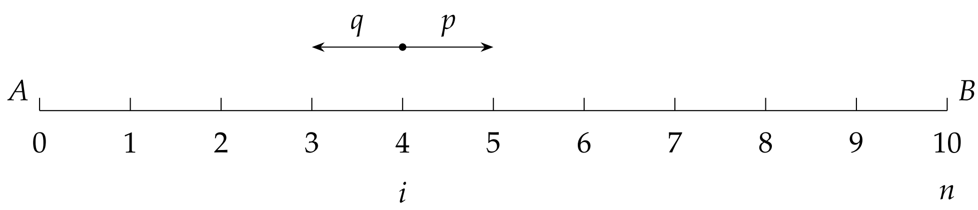
The programs are written in the Python 3 language and use the `matplotlib` library to generate the graphs. You need to install Python (`https://www.python.org/downloads/`) although a knowledge of Python programming is not necessary to run the simulations.

Parameters directly related to the problems, such as the probability of success, can be modified interactively. Others, related to the simulation, such as the number of steps in a simulation, are defined in a file `configuration.py`. The code that uses `matplotlib` also appears in a separate file.

2 Gambler's ruin

2.1 Gambler's ruin with no draw

Two players A and B compete in a contest. There is an initial finite capital of n units: A has i and B has $n-i$. They repeatedly play a game where the probability that A wins is p and the probability that B wins is $q = 1 - p$. The loser gives one unit to the winner. When one player has all n units the contest is terminated and that player is declared the winner.



- Given initial parameters (p, n, i), what is the probability that A wins?
- What is the expected duration of the game?

Given (p, n, i) the probability that A wins the contest is:

$$P_A(p, n, i) = \left(\frac{1 - r^i}{1 - r^n}\right),$$

where $r = q/p$. By symmetry, the probability that B wins is:

$$P_B(p, n, i) = \left(\frac{1 - (1/r)^{n-i}}{1 - (1/r)^n}\right).$$

For $p \neq 1/2$ the expected duration of the contest is:

$$E_{duration}(p,n,i) = \frac{1}{q-p}\left(i - n\frac{1-r^k}{1-r^n}\right) ,$$

while for $p = 1/2$ the expected duration is:

$$E_{duration}(p,n,i) = i(n-1) .$$

Of course the duration does not depend on which player wins. If A wins, the contest terminates for B and conversely.

You can run a simulation more than once with the saved parameters, enter new parameters, or run a sequence of simulations for a range of probabilities or initial values. Here is one output:

```
Probability = 0.45, capital = 20, initial = 8
Wins = 789, losses = 9211, limits exceeded = 0
Proportion of wins     = 0.0789
Probability of winning = 0.0732
Average duration  = 65
Expected duration = 65
```

A graph of the proportion of wins and the histogram of the durations are shown in Figures 1, 2. The vertical lines in the histograms are the average durations.

Sources: [7, Chapter 2], [6, 2, Problems 35, 36], [8, Example 4m], [3, Example 2.7.3], [4, Section 16.16], [5, Section 20.1].

2.2 Gambler's ruin with draw

The problem can be modified so that the probability that A wins is $p \leq 1/2$, the probability that B wins is also p, and the probability that there is a draw and no units are exchanged is $1-2p$. The probability that A wins is now $(n-i)/n$ and the expected duration of the game is:

$$\frac{i(n-i)}{2p} .$$

Here is the result of a simulation:

```
Probability = 0.35, capital = 20, initial = 10
Wins = 4985, losses = 5014, limits exceeded = 1
Proportion of wins     = 0.4985
Probability of winning = 0.5000
Average duration  = 142
Expected duration = 142
```

The graphs are the same as for gambler's ruin with no draw.

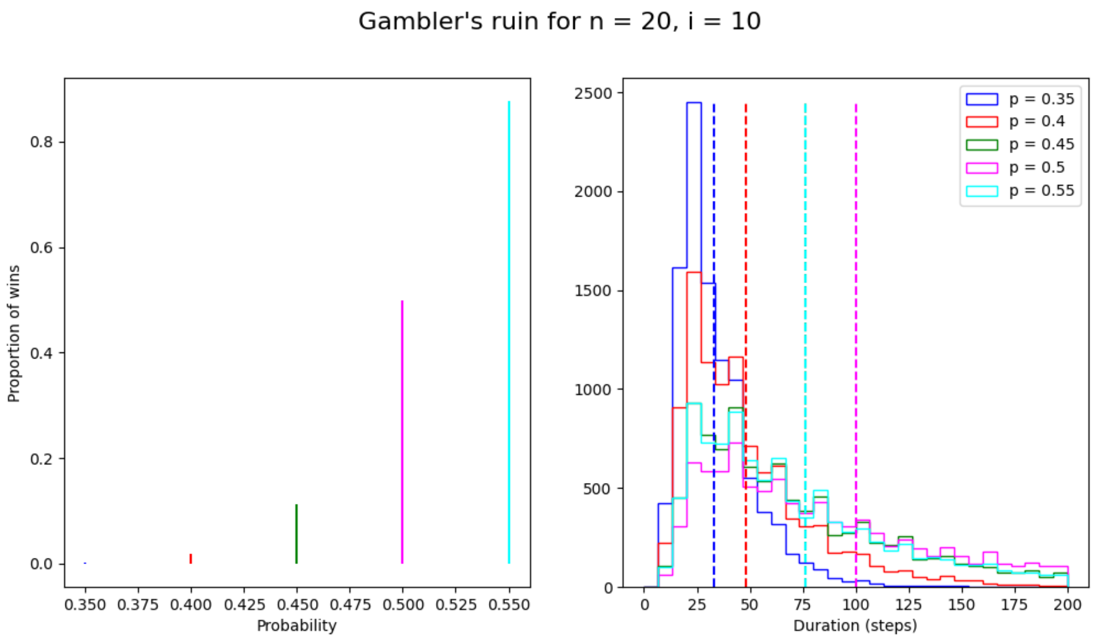


Figure 1: Proportion of wins and histogram for $n = 20, i = 10$ and multiple probabilities

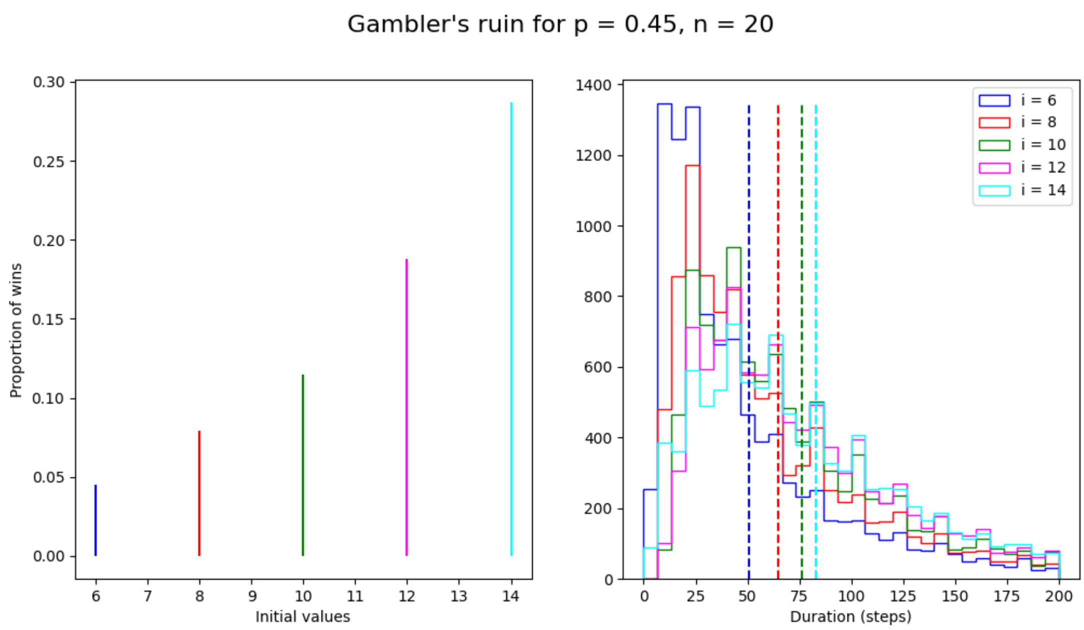


Figure 2: Proportion of wins and histogram for $p = 0.45, n = 20$ and multiple initial values

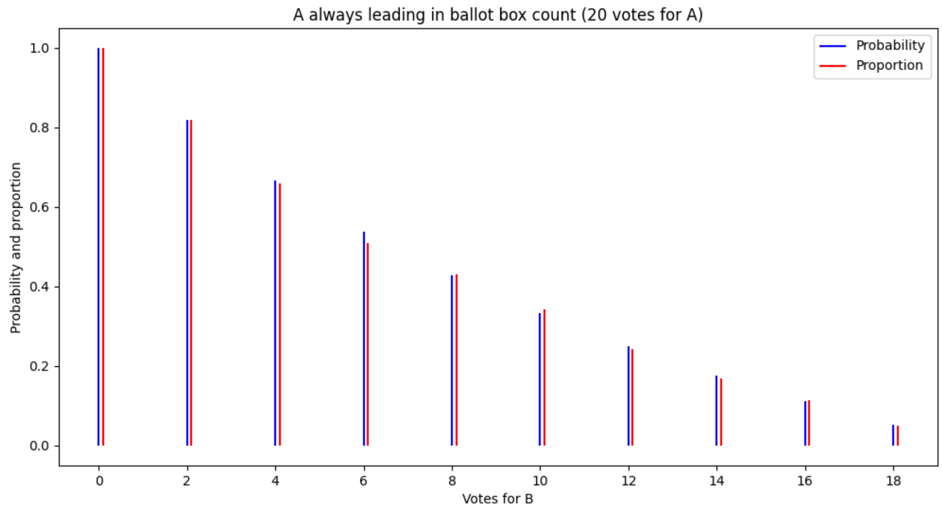


Figure 3: Probability and proportion of A always leading

3 The ballot box

In an election there are two candidates A and B. A receives a votes and B receives b votes where $a > b$. The votes are counted one-by-one and the running totals $(a_i, b_i), 1 \leq i \leq a+b$ are updated as each vote is counted. What is the probability that $a_i > b_i$ always holds?

The probability is:

$$P(A \text{ is always leading}) = \frac{a-b}{a+b}.$$

The result of a simulation is:

```
For a = 20, b = 18:
Probability of A always leading = 0.0526
Proportion  of A always leading = 0.0510
```

A graph of the probability and the proportion of simulations that A is always leading is shown in Figure 3. As b gets closer to a the probability decreases since more and more votes for B are available to be counted.

Sources: [6, 2, Problem 22], [4, Section 16.3]

4 Random walk

4.1 One-dimensional random walk

A particle is placed at the origin of the x-axis. It repeatedly takes steps: right with probability p and left with probability $q = 1 - p$.

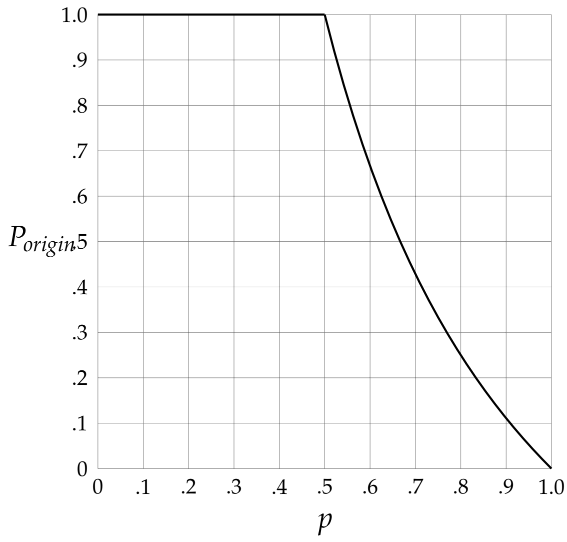
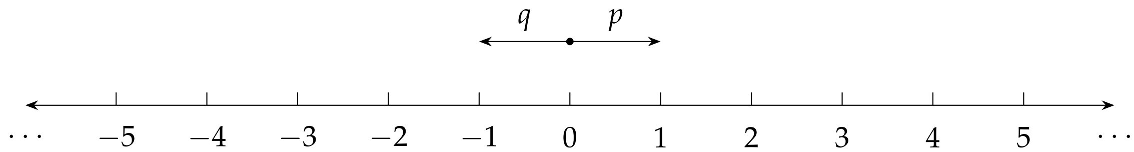


Figure 4: Graph of P_{origin}



- What is the probability that the particle will return to the origin?
- What is the expected duration until the particle returns to the origin?

By symmetry let the first step be to the right. The particle can only return to the origin after an even number of steps. Assume that $p = 1/2$. Let S_{2m} be the position of the particle after $2m$ steps. Then:

$$P(S_{2m} = 0) = \binom{2m}{m} \frac{1}{2^{2m}},$$

which by Stirling's formula is approximately equal to $1/\sqrt{\pi m}$. It can now be proved that the probability of a return to the origin is 1.

For $p \leq 1/2$, P_{origin}, the probability of a return to the origin, is 1 and for $p \geq 1/2$ the probability is (Figure 4):

$$P_{origin} = \frac{q}{p} = \frac{1-p}{p}.$$

E_{origin}, the expected duration until the first return to the origin, is infinite for $p \geq 1/2$ while for $p < 1/2$ it is:

$$E_{origin} = \frac{1}{q-p} = \frac{1}{1-2p}.$$

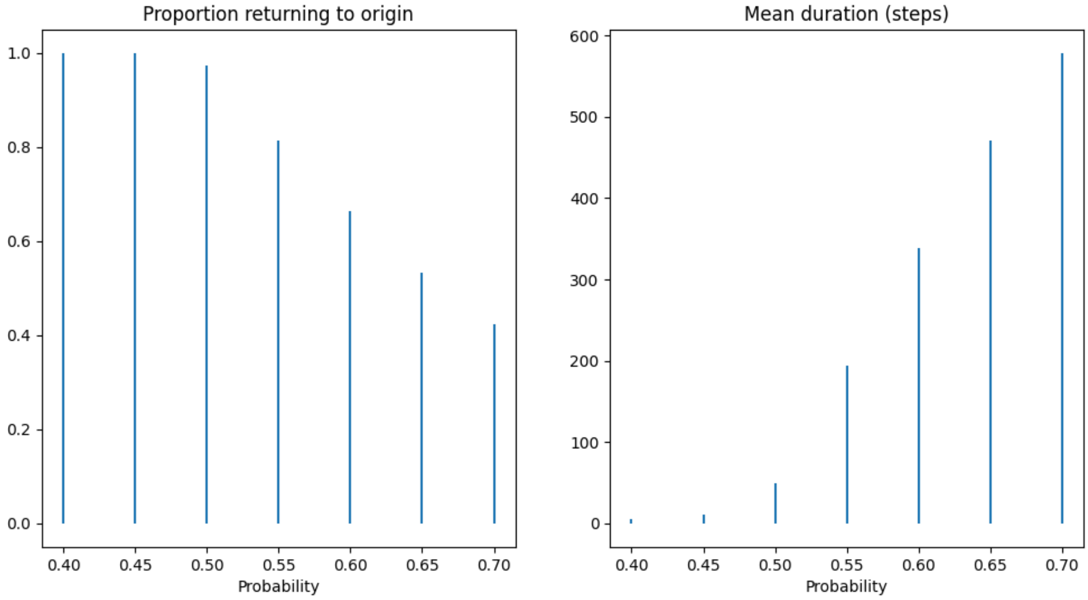


Figure 5: Proportion of returns to origin and and mean durations for multiple probabilities

You can run a simulation more than once with the saved parameters, enter new parameters, or run a sequence of simulations for a range of probabilities or limits. Here is one output:

```
Probability = 0.50, step limit   = 1000
Proportion returning to origin   = 0.977
Probability of return to origin  = 1.000
Proportion reaching limit        = 0.023
Mean duration (steps)            = 49
Expected duration (steps)        = infinity
```

The proportion of wins in the simulation are very close to the theoretical probability, but the mean duration is far from infinite because the step limit was too small. The proportion of wins and the mean durations are shown in Figures 5 and 6.

Sources: [7, Chapter 3], [3, Example 3.7.2], [4].

4.2 Two-dimensional random walk

In a two-dimensional random walk a step of the particle consists of one step left or right on the x-axis with probability $1/2$ and simultaneously one step up or down on the y-axis also with probability $1/2$ (Figure 7).

The probability 1 the particle will return to the origin but the expected duration is infinite! Therefore, when you run the simulation with any reasonable limit on the number of steps,

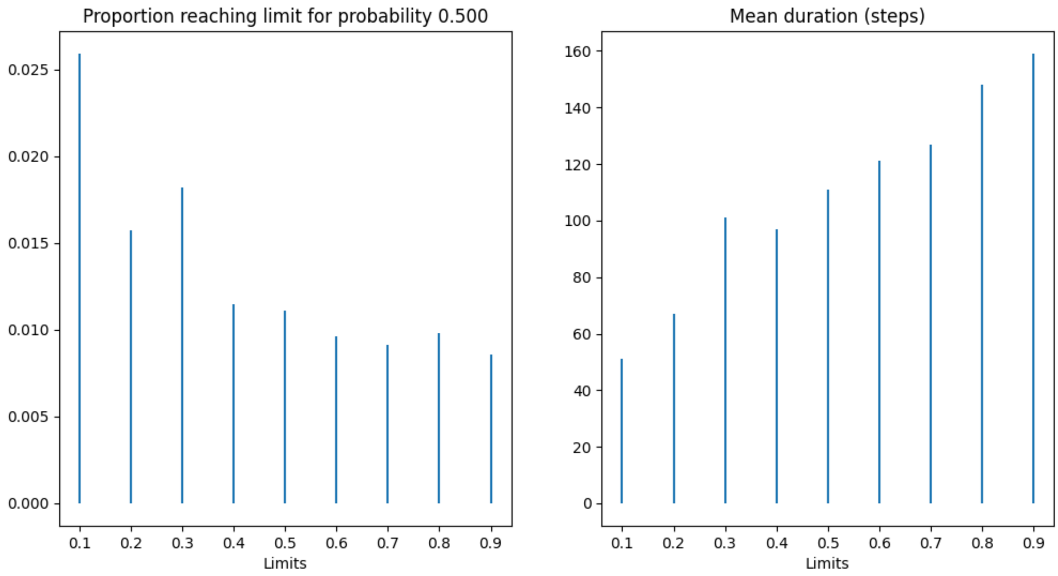


Figure 6: Proportion of returns to origin and mean durations for multiple limits

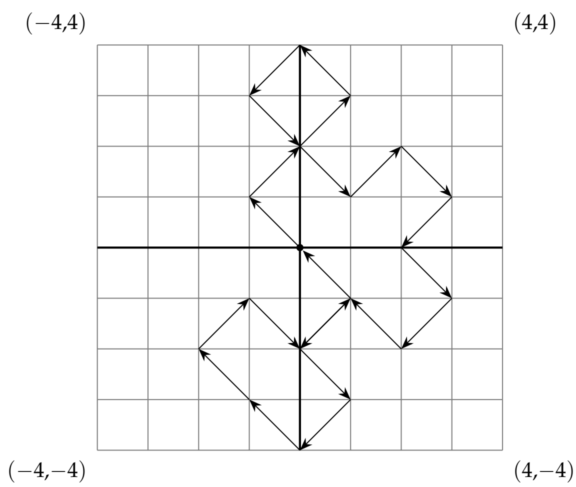


Figure 7: A 22-step two-dimensional random walk

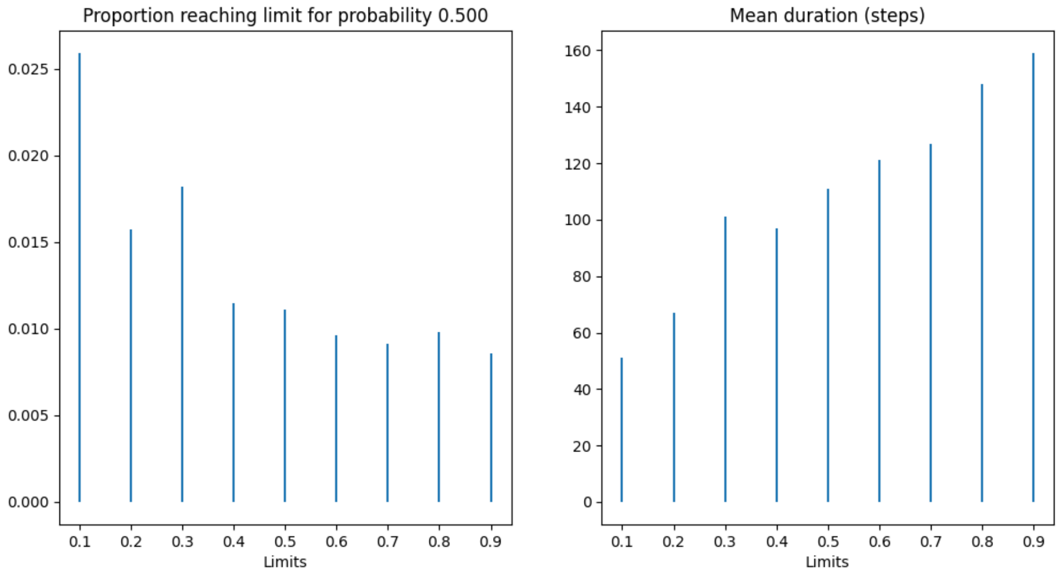


Figure 8: Proportion of returns to origin and and mean durations for multiple limits

the proportion of returns to the origin will be much less than 1 and the mean duration will be quite large:

```
Limit                          = 100000
Proportion returning to origin = 0.777
Proportion reaching limit      = 0.223
Mean duration (steps)          = 24133
```

Figure 8 shows the proportion of returns to the origin and the mean durations for a range of fractions of the limit.

Source: [6, 2, Problem 51].

4.3 Three-dimensional random walk

The three-dimensional random walk adds a simultaneous step along the z-axis with probability $1/2$. The probability of a return to the origin is only 0.2379 so the simulations will show a large number of simulations reaching the limit:

```
Limit                          = 100000
Proportion returning to origin = 0.370
Proportion reaching limit      = 0.630
Mean duration (steps)          = 63518
```

Source: [6, 2, Problem 52].

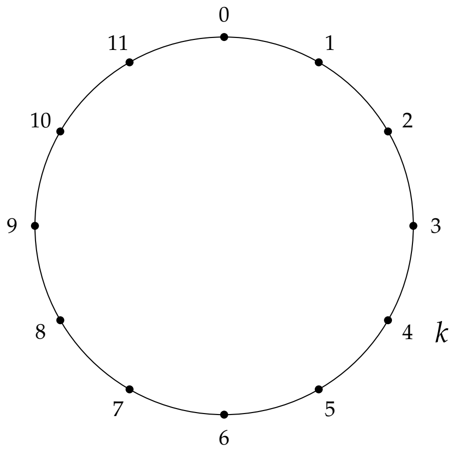


Figure 9: Random walk in a circle

4.4 Random walk in a circle

Consider a circle with n equally-spaced points around its circumference (Figure 9). A random walk starts at point 0 and proceeds to the next point or the previous point both with probability $1/2$.

- With probability 1 the all points will be visited.
- With probability $1/n$ a designated point k will be the last point visited.

The random walk will eventually visit all points, but it make take a very large number of steps. (Try to run the simulation for 100 points!) Here is an output:

```
For 50 points, 1000 simulations, at most 3000 steps:
  Designated point visited 22 times
    Probability = 0.020, proportion = 0.022
  Did not visit all points 26 times
```

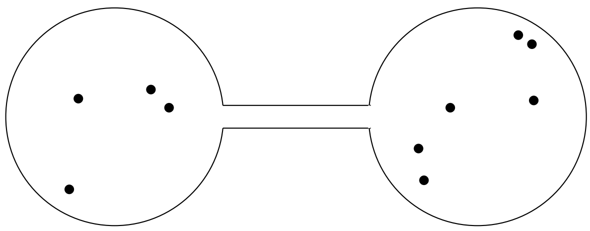
The histogram of steps visited is shown Figure 10.

Source: [1, Section 3].

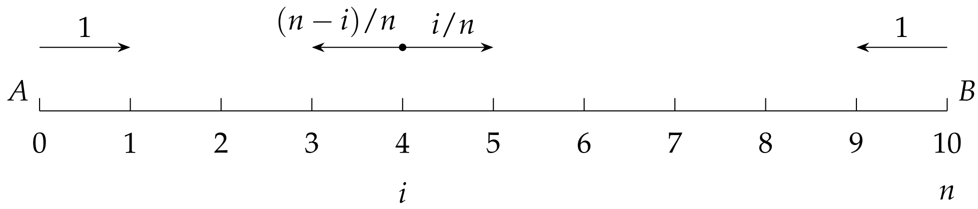
5 The Ehrenfest model

The Ehrenfest model models the diffusion of particles between two containers. In the following diagram there are 4 particles in the left container and 6 in the right container.

Figure 10: Histogram of steps until all points are visited



Repeatedly choose a particle at random with uniform distribution and move it to the other container. If there are i particles in the left container then the probability of choosing a particle from the left container is i/n and the probability of choosing a particle from the right container is $(n-i)/n$. If one container is empty the next particle must be chosen from the other container.



The problem is similar to the gambler's ruin except that the process never ends and the probability of a left or right step changes with each step.

The process is a Markov chain which eventually reaches a *stationary distribution*:

$$s_i = \binom{n}{i} \left(\frac{1}{n}\right)^n ,$$

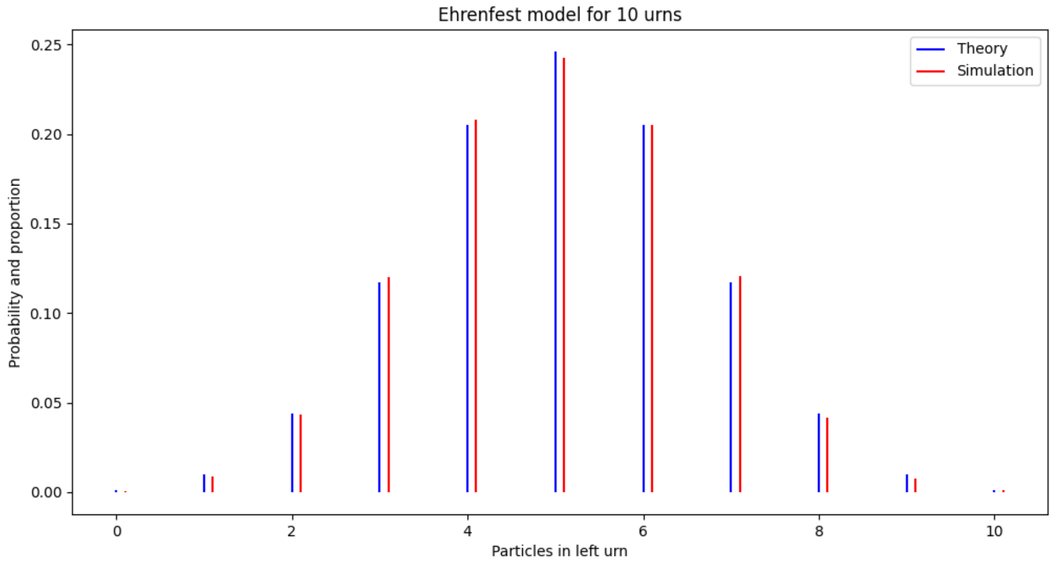


Figure 11: Stationary distribution for the Ehrenfest model

where s_i is the proportion of time that the particle is at the i'th position.

Here is an output of the simulation:

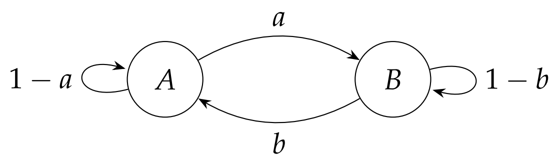
```
Total particles in urns = 10
Theoretical stationary distribution
[0.001 0.01  0.044 0.117 0.205 0.246 0.205 0.117 0.044 0.01  0.001]
Simulation stationary distribution
[0.001 0.009 0.044 0.12  0.208 0.243 0.205 0.121 0.042 0.008 0.001]
```

A graph of these distributions is shown in Figure 11; the theoretical distribution and the result of simulation are so close together that the lines are slightly offset.

Sources: [3, Example 11.4.6], [7, Section 4.3].

6 The two-state process

The two-state process is similar to the Ehrenfest model in that the probabilities at each step are different and we are interested in the stationary probability distribution of the unbounded process. There are two states A, B. In state A the process transitions to B with probability a and remains in A with probability $1 - a$. Similarly, the probability of a transition from B to A is b and the probability of remaining in B is $1 - b$.



The stationary distribution, that is, the proportion of visits to A and to B is:

$$\left[\frac{b}{a+b}, \frac{a}{a+b}\right].$$

Here is an output of a simulation:

```
Probabilities:  a = 0.500, b = 0.333
Theoretical stationary distribution: A = 0.400, B = 0.600
Simulation  stationary distribution: A = 0.402, B = 0.598
```

When $a + b = 1$ the probability of being at A is b and the probability of being at B is a:

```
Probabilities:  a = 0.333, b = 0.667
Theoretical stationary distribution: A = 0.667, B = 0.333
Simulation  stationary distribution: A = 0.674, B = 0.326
```

You can enter a required proportion p of visits to B and any probability $0 < a < p$. The proportion will be achieved for:

$$b = \frac{a(1-p)}{p},$$

as shown in the following simulation where we entered $p = 0.8, a = 0.6$:

```
Probabilities:  a = 0.600, b = 0.150, proportion = 0.800
Theoretical stationary distribution: A = 0.200, B = 0.800
Simulation  stationary distribution: A = 0.194, B = 0.806
```

A plot of the stationary distribution for A and B for fixed a and various values of b is shown in Figure 12.

Sources: [7, Section 4.5], [3, Example 11.1.3]

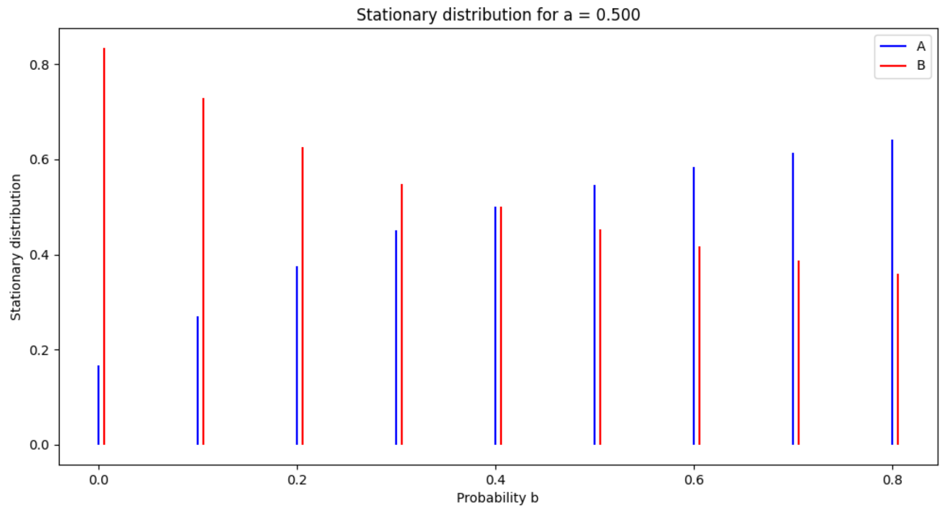


Figure 12: Stationary distribution for the two-state model

7 A maze

The Markov chain in Figure 13 represents a maze since all the transitions are bidirectional.

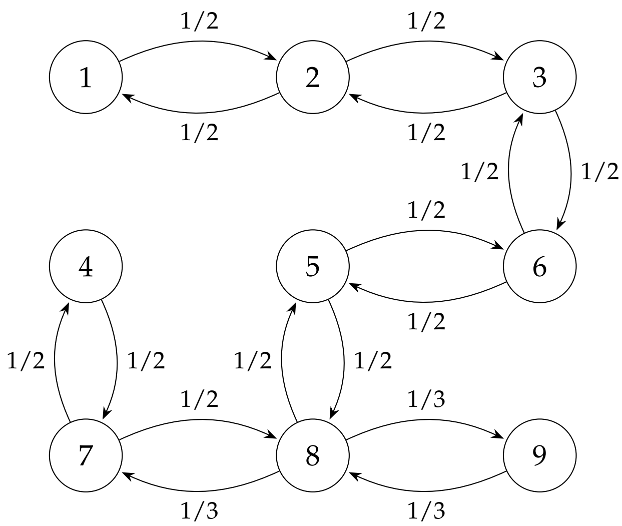


Figure 13: The states and transitions of a maze

7.1 Average number of returns

What is the average number of steps until the first return to state *i* starting in state *j*?

Privault [7, Sections 5.3] computes the expected number of returns to state 0 starting in any other state and the average number of steps in the simulation is very close to to the expected number of returns:

```
Expected steps to return to 0 = [16, 15, 28, 59, 48, 39, 58, 55, 56]
Average  steps to return to 0 = [14, 15, 28, 59, 46, 39, 59, 55, 54]
```

The simulation of the average number of returns to any state *i* from any state *j* gives:

```
Average  steps to return to 0 = [14, 15, 28, 59, 46, 39, 59, 55, 54]
Average  steps to return to 1 = [ 1,  8, 14, 43, 31, 24, 42, 40, 40]
Average  steps to return to 2 = [ 3,  2,  8, 31, 20, 10, 29, 27, 30]
Average  steps to return to 3 = [52, 53, 50, 15, 37, 44, 14, 29, 28]
Average  steps to return to 5 = [ 8,  8,  5, 20,  8,  7, 18, 15, 16]
Average  steps to return to 6 = [37, 37, 35,  1, 21, 29,  8, 13, 14]
Average  steps to return to 7 = [25, 23, 21,  4,  9, 16,  3,  5,  1]
Average  steps to return to 8 = [38, 37, 37, 19, 25, 30, 18, 14, 16]
```

Privault [7, Sections 7.2] also computes the expected number of returns to state *i* from state *i* and the averages (the main diagonal of the above matrix) are very close:

```
Expected steps to return to i from i = [16,  8,  8, 16,  8,  8,  8,  5, 16]
Average  steps to return to i from i = [14,  8,  8, 15,  7,  7,  8,  5, 16]
```

Source: [7, Section 5.3].

7.2 Stationary distribution

Let π be a probability distribution of the initial state of a Markov chain. Take one step according to the transition matrix. If the probability distribution is still π, it is the *stationary distribution* of the chain. Clearly, the distribution will remain the same no matter how many steps are taken.

For each step in the simulation the initial state is randomly selected according to the distribution and then the transition matrix is used to compute the next state. A count of the these states is maintained and used to obtain a simulated distribution.

Privault computes the stationary distribution of the maze and the result of the simulation is very close:

```
Stationary distribution:
[0.0625, 0.1250, 0.1250, 0.0625, 0.1250, 0.1250, 0.1250, 0.1875, 0.0625]
Distribution after first step:
[0.0626, 0.1277, 0.1269, 0.0612, 0.1248, 0.1261, 0.1240, 0.1883, 0.0584]
```

Source: [7, Section 7.2].

8 Branching processes

A branching process starts out with a single state $\{s_{00}\}$. From that state a random number of level 1 new states $\{s_{10}, s_{11}, \ldots, s\}$ is created according to some distribution and the initial state is deleted. Recursively and according to the same distribution, a new level is created with a random number of successors to each state at the current level.

8.1 Mean population

What is μ_n the mean number of states X_n at level n? Privault proves that:

$$\mu_n = E(X_n \mid X_0 = 1) = E(X_1 \mid X_0 = 1)^n = \mu_1^n .$$

The simulation can be run for three distributions:

- Throwing a fair die, expectation 3.5.
- Throwing a fair die until a six is obtained, expectation 6.
- Collecting coupons until all five coupons are collected, expectation:

  $$E(\text{all } n \text{ numbers}) = n\left(\frac{1}{n} + \frac{1}{n-1} + \cdots \frac{1}{2} + \frac{1}{1}\right) = nH_n \approx n\left(\ln n + \frac{1}{2n} + 0.5772\right),$$

  where H_n is the n-th harmonic number.

```
Throwing a fair die, levels = 4
Expectation of mean population size: 150
Simulatation   mean population size: 157

Throwing a die until a six appears, levels = 4
Expectation of mean population size: 1296
Simulatation   mean population size: 1228

Coupon collector (five coupons), levels = 4
Expectation of mean population size: 17087
Simulatation   mean population size: 17550
```

Figure 14 shows a histogram of the population of the coupon collector at level 3, together with the expectation and the mean.

Source: [7, Section 8.2], [6, 2, Problems 4, 14].

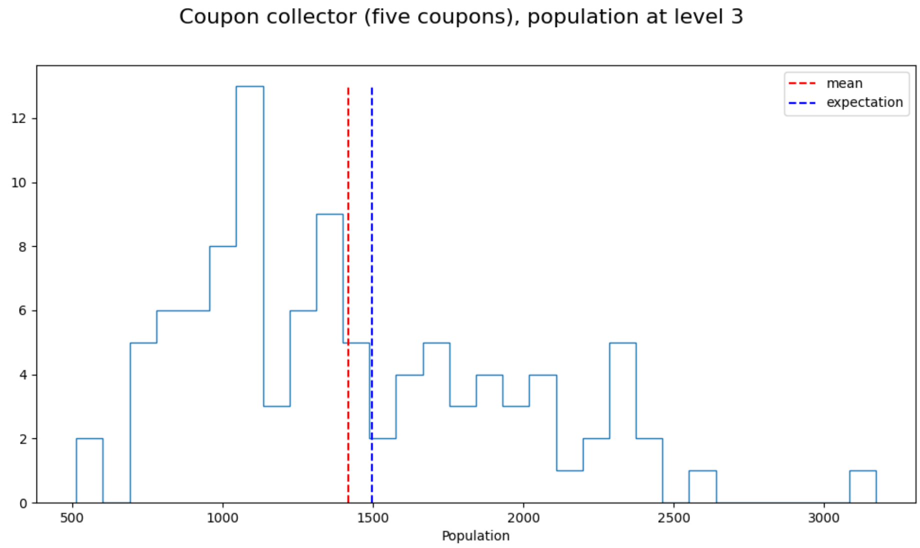


Figure 14: Histogram of the populations

8.2 Extinction probability

If a branching process starts with one state, what is $P(E)$ the probability that it will become *extinct*, that is, that the population will become zero. Privault [7, Section 8.3] works out two simple cases:

- $P(Y_1 = 1 \mid X_1 = 1) = p$, $P(Y_1 = 0|X_1 = 1) = 1 - p$. Each new level consists of one state with probability p, otherwise extinction occurs.
- $P(Y_1 = 2 \mid X_1 = 1) = p$, $P(Y_1 = 0|X_1 = 1) = 1 - p$. At each level a state is replaced by two states with probability p, otherwise it has no successors.

The probability for the first case is not interesting: $P(E) = 0$ if $p = 1$, otherwise $P(E) = 1$. For the second case: if $1 - p < p$[1] then $P(E) = (1 - p)/p$, otherwise $P(E) = 1$.

The choice of the distribution, and theprobability and number of levels can be set interactively. Here is the result of a simulation:

```
Distribution is P(Y_1=1 | X_1=1)
Probability = 0.500, levels = 8
Probability of extinction  = 1.000
Proportion  of extinctions = 0.990
```

The number of levels should be increased until the proportion approaches the probability. For the second case, the simulation can be run multiple times for various probabilities and the results ploted (Figure 15).

[1] This is the same as $p > 1/2$.

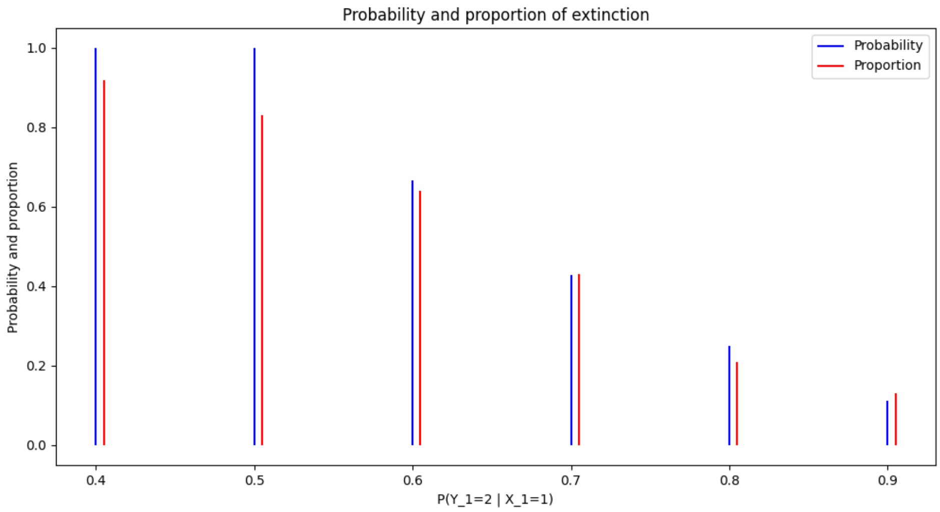


Figure 15: Probability and proportion of extinction

Source: [7, Section 8.3].

References

[1] Sven Erick Alm. Simple random walk. `http://www2.math.uu.se/~sea/kurser/stokprocmn1/slumpvandring_eng.pdf`, 2006.

[2] Moti Ben-Ari. Mosteller's challenging problems in probability. `https://github.com/motib/probability-mosteller`, 2022.

[3] Joseph K. Blizstein and Jessica Hwang. *Introduction to Probability (Second Edition)*. CRC Press, 2019.

[4] K.C. Border. Lecture 16: Simple random walk. `http://www.math.caltech.edu/~2016-17/2term/ma003/Notes/Lecture16.pdf`, 2017.

[5] Eric Lehman, F. Thomson Leighton, and Albert R. Meyer. *Mathematics for Computer Science*. MIT Open Courseware, 2015. `https://ocw.mit.edu/courses/6-042j-mathematics-for-computer-science-spring-2015/pages/readings/`.

[6] Frederick Mosteller. *Fifty Challenging Problems in Probability with Solutions*. Dover, 1965.

[7] Nicolas Privault. *Understanding Markov Chains: Examples and Applications (Second Edition)*. Springer, 2018.

[8] Sheldon Ross. *A First Course in Probability (Tenth Edition)*. Pearson, 2019.