

Math 365 Polya's Urn

Submit an R script and/or handwritten submission by next Wednesday.

The objective of this lab is to explore an example of a martingale called *Polya's urn*.

1 Description of Polya's Urn

Suppose you have an urn containing one red ball and one green ball. You draw one at random; if the ball is red, put it back in the urn with an additional red ball (so urn now has 2 red and 1 green), otherwise put it back and add a green ball (so urn now has 1 red and 2 green). Repeat this procedure and let the random variable X_n be the number of red balls in the urn after n draws. Note that $X_0 = 1$.

The random variables X_0, X_1, X_2, \dots form a Markov chain with transition probabilities

$$\begin{aligned}\mathbb{P}\{X_{n+1} = k \mid X_n = k\} &= 1 - \frac{k}{n+2} \\ \mathbb{P}\{X_{n+1} = k+1 \mid X_n = k\} &= \frac{k}{n+2}\end{aligned}$$

Let $M_n = \frac{1}{n+2}X_n$, the proportion of red balls in the urn after n draws. To prove that M_n is a martingale with respect to X_n , it's sufficient to show that $E(M_{n+1} \mid X_n) = M_n$. Check the calculations below to make sure they make sense to you:

$$E(X_{n+1} \mid X_n)(k) = k\mathbb{P}\{X_{n+1} = k \mid X_n = k\} + (k+1)\mathbb{P}\{X_{n+1} = k+1 \mid X_n = k\} = \frac{n+3}{n+2}k,$$

so

$$E(M_{n+1} \mid X_n) = \frac{1}{n+3}E(X_{n+1} \mid X_n) = \frac{1}{n+3} \cdot \frac{n+3}{n+2}X_n = \frac{1}{n+2}X_n = M_n.$$

2 Simulation of Polya's urn

The following R code simulates 998 draws, until the urn contains 1000 balls. The simulation is run 2000 times, so you can examine the distribution of M_{998} at the end:

```
nStop <- 998
nSim <- 2000
propRed <- matrix(0,nSim,1)
for (n in 1:nSim) {

  nRed <- 1
  nGreen <- 1
  for (k in 1:nStop) {
    drawnball <- sample(0:1,size=1,prob=c(nRed,nGreen)/(nRed+nGreen))
    nRed <- nRed+(1-drawnball) # drawnball=0 if red, 1 if green
    nGreen <- nGreen+drawnball
```

```
}  
propRed[n] <- nRed/(nRed+nGreen)  
  
}  
hist(propRed)
```

Exercise 1 Conjecture what the distribution of M_n is in general, based on the simulation results.

Exercise 2 Prove using induction on n that

$$\mathbb{P}\left\{M_n = \frac{k}{n+2}\right\} = \frac{1}{n+1} \text{ for } k = 1, 2, \dots, n+1.$$

That is, the distribution is uniform. Hint: as we often do with Markov chain calculations, look one step back in time at the possible previous values.

3 Another simulation of Polya's urn

Exercise 3 Revise the simulation code to run 100 simulations that each do 1,998 draws (so total of 2,000 balls when done). Store the proportion of red balls after 998 draws and also after 1,998 draws. How do these numbers compare for each simulation? Plot M_{1998} versus M_{998} from the 100 simulations. What is happening?