Stevens Institute of Technology

SYS-611 Homework #5

Due Mar. 31 2021

Submit the following using the online submission system: 1) Cover sheet with name, date, and collaborators, 2) Written responses in PDF format, 3) All work (e.g. .xlsx or .py files).

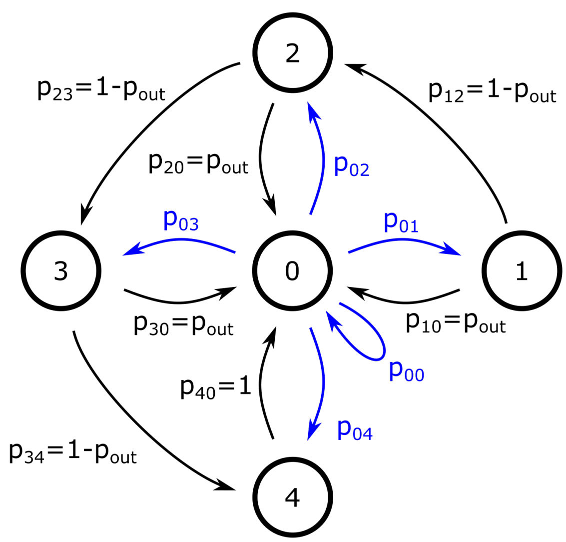
5.1 Ty Cobb's Markov Model [13 points]

Ty Cobb played professional baseball for the Detroit Tigers and Philadelphia Athletics from 1905 to 1928. He holds the record for highest career batting average and also set the single-season record in 1915 for stolen bases (96) which was not surpassed for almost 50 years. In the 1915 season Cobb faced a pitcher precisely 700 times, recording the following outcomes[1]:

- Reached first base on a hit, walk, or hit-by-pitch 289 times
- Reached second base on a hit (double) 31 times
- Reached third base on a hit (triple) 13 times
- Reached home plate on a hit (home run!) 3 times
- Recorded an out (strike out, fly out, ground out, or sacrifice) 364 times

Furthermore, during the season, he was able to advance bases (e.g. from first to second, etc.) on 96 of 134 attempts. For the other 38 attempts he was caught and called "out".

Consider a Markov model with five states: (0): Out/Bench, (1): First base, (2), Second base, (3), Third base, (4) Home plate. Assume Cobb plays alone and moves by stealing bases or getting caught. If he reaches home plate, he immediately returns to the bench.



[1] Data is courtesy Sean Lahman's Baseball Database, `http://www.seanlahman.com/baseball-archive/statistics/`, available under a Creative Commons Attribution-ShareAlike 3.0 Unported License

(a) 3 PTS Use the data provided above to estimate the state transition matrix $P = [p_{ij}]$.

(i) Use batting data to find state transition probabilities of recording an out (p_{00}), reaching first base (p_{01}), second base (p_{02}), third base (p_{03}), or home plate (p_{04}).

(ii) Use stolen base data to find state transition probabilities of advancing bases: $p_{12} = p_{23} = p_{34} = 1 - p_{\text{out}}$ or being caught and called out: $p_{10} = p_{20} = p_{30} = p_{\text{out}}$.

(iii) Fill in values for the state transition matrix $P = [p_{ij}]$. (Note: many are 0).

(b) 5 PTS Complete a discrete process generator to return $q(t+1)$ for each of the following cases provided a random (0,1) sample r:

(i) If Cobb is on the bench (i.e. current state is $q(t) = 0$).

$$q(t+1|q(t)=0) = \begin{cases} 0 & \text{if } r \leq \underline{\hspace{3cm}} \\ 1 & \text{if } \underline{\hspace{3cm}} < r \leq \underline{\hspace{3cm}} \\ 2 & \text{if } \underline{\hspace{3cm}} < r \leq \underline{\hspace{3cm}} \\ 3 & \text{if } \underline{\hspace{3cm}} < r \leq \underline{\hspace{3cm}} \\ 4 & \text{if } \underline{\hspace{3cm}} < r \leq \underline{\hspace{3cm}} \end{cases}$$

(ii) If Cobb is on a base (i.e. current state is $q(t) \in \{1, 2, 3\}$).

$$q(t+1|q(t) \in \{1,2,3\}) = \begin{cases} 0 & \text{if } r \leq \underline{\hspace{3cm}} \\ q(t)+1 & \text{if } \underline{\hspace{3cm}} < r \leq \underline{\hspace{3cm}} \end{cases}$$

(iii) If Cobb is on home plate (current state is $q(t) = 4$). (Hint: a simple case).

$$q(t+1|q(t)=4) = \underline{\hspace{3cm}}$$

(c) 2 PTS **By hand**, complete the following manual simulation for provided random (0,1) samples r using an initial state of $q(0) = 0$.

t	r	$q(t)$	$q(t+1)$
0	0.765	0	
1	0.557		
2	0.347		
3	0.098		
4	0.039		
5	0.132		

(d) 3 PTS **By computer**, simulate $N = 1000$ state transitions using a stream of random (0,1) samples. Estimate the stationary stochastic distribution π (steady-state probability of each state) by finding the relative frequency Cobb is in each state.

5.2 Cashier Wanted at Café Java [12 points]

The manager of Café Java is looking to hire a new cashier to handle $\lambda = 30$ customers per hour (0.5 per minute). The decision is narrowed down to two candidates. Candidate A takes an average of $1/\mu_A = 1.5$ minutes to prepare a coffee (i.e., $\mu_A = 2/3$ per minute). Candidate B takes an average of $1/\mu_B = 1.2$ minutes to prepare a coffee (i.e., $\mu_B = 5/6$ per minute). Assume customer inter-arrival and service times are exponentially distributed.

(a) 2 PTS Using queuing theory:

- (i) Find the utilization ratio ρ for each candidate (A and B).
- (ii) Find the average waiting time $\overline{W}$ for each candidate (A and B).

(b) 2 PTS **By hand**, simulate the following state variables for each event i:

- $t(i)$: time
- $q(i)$: number of customers in the queuing system
- $t_{\text{arrival}}(i)$: sampled inter-arrival duration until next customer
- $t_{\text{service}}(i)$: sampled service duration

derived state variable $\Delta t(i)$ to measure the duration of event i

$$\Delta t(i) = \begin{cases} t_{\text{arrival}}(i) & \text{if } q(i) = 0 \text{ or } t_{\text{arrival}}(i) < t_{\text{service}}(i) \\ t_{\text{service}}(i) & \text{otherwise} \end{cases}$$

and state transition functions:

$$q(i+1) = \begin{cases} q(i) + 1 & \text{if } \Delta t(i) = t_{\text{arrival}}(i) \\ q(i) - 1 & \text{otherwise} \end{cases}$$

$$t(i+1) = t(i) + \Delta t(i)$$

for the following initial state and sampled times:

i	$t(i)$	$q(i)$	$t_{\text{arrival}}(i)$	$t_{\text{service}}(i)$	$\Delta t(i)$	$q(i+1)$	$t(i+1)$
0	0.0	0	2.97	1.19			
1			2.52	2.67			
2			3.07	0.13			
3			2.42	2.03			
4			2.56	1.25			

(c) 2 PTS **By hand**, add derived state variables to measure how long the cashier is busy

$$b(i) = \begin{cases} \Delta t(i) & \text{if } q(i) > 0 \\ 0 & \text{otherwise} \end{cases}$$

the number of new customer arrivals

$$c(i) = \begin{cases} 1 & \text{if } \Delta t(i) = t_{\text{arrival}}(i) \\ 0 & \text{otherwise} \end{cases}$$

and the total waiting time incurred during an event

$$w(i) = q(i) \cdot \Delta t(i)$$

to complete the additional columns in the table started in part (b):

i	...	$b(i)$	$c(i)$	$w(i)$
0	...			
1	...			
2	...			
3	...			
4	...			

(d) 4 PTS **By computer**, implement the simulation model in a computational modeling tool with process generators for t_{arrival} and t_{service}, simulate at least $N = 1000$ events, and report the utilization ratio

$$\rho = \frac{\sum_{i=0}^{N} b(i)}{\sum_{i=0}^{N} \Delta t(i)}$$

and average waiting time

$$\overline{W} = \frac{\sum_{i=0}^{N} w(i)}{\sum_{i=0}^{N} c(i)}$$

for each candidate (A and B). *Note: A and B require different generators for $t_{service}$.*

(e) 1 PT Compare results from (a) and (d) and comment on any observed differences. How could you get a more accurate estimate of the steady-state utilization ratio and average waiting time from simulation?

(f) 1 PT What causes simulation results to be biased for a small number of events (N)?