Mathematical Modelling: Using Markov chains to model bathroom queues

Johnny Wong[1]

You come home from your morning run all sweaty and ready for a shower. As you approach the bathroom, you see the locked door and roll your eyes. Standing in your sweat soaked singlet, your little brother pokes his head out of his bedroom, chucks a deodorant at you before telling you he had already called dibs on the bathroom after you sister is finished. You throw the deodorant back at him and curse your house for having so few bathrooms.

Everyone has had to wait to use the bathroom at some point in their lives. It's common sense that the fewer bathrooms or more people there are, the longer you'd have to wait. But can we mathematically calculate how long we can expect to wait every day? One approach is with a technique called Markov chains.

For the rest of the article, let's consider a household with one bathroom between 4 people.

Markov chains

To set up a Markov chain, you need several things: *states* that represent the possible scenarios, *probabilities* of moving from state to state throughout some measure of *time*. We will unpack these 3 things in more detail below.

States

A state is something that can describe the situation of interest at different points in time.

For our bathroom queueing problem, we are interested in how many people are using, or wanting to use, the bathroom at any point in time. So how many states do we need? Well with 4 people, there can be from 0 to 4 people wanting to use the bathroom, meaning we need 5 states.

Let's label each state $0, \cdots, 5$, where state 0 means no one is wanting to use the bathroom. State 4 means four people want to use the bathroom, since there is only one bathroom available, this means 1 person is using it and there are 3 queuing up.

Time

Time can be either discrete or continuous. With this scenario, it is most appropriate to use continuous time as people can enter and leave the bathroom at any point in time.

---

[1]Johnny Wong is a recent graduate of UNSW, Australia (`johnny.c.wong@unswalumni.com`)

Let's represent the state at time t as X_t

Probability

How does the system jump from state to state? A fundamental assumption of Markov chains is that the probability of moving from any state to the next is not impacted by any past information, and only depends on the current state.

In continuous time, these probabilities are represented as transition rates of moving from state to state.

Transition rates What is a transition rate? Roughly speaking, it is the probability per time unit that the system makes a transition from one state to the other. if $q_{i,j}$ represents the transition rate from state i to j where $i \neq j$, and if the current state is i, the probability that the state is j after time h has passed is roughly $h \times q_{i,j}$ for small h

$$Pr(X_{t+h} = j | X_t = i) \approx h \times q_{i,j} \qquad \text{for small } h$$

More precisely,

$$q_{i,j} = \lim_{h \to 0} \frac{Pr(X_{t+h} = j | X_t = i)}{h}$$

Of course, when we specify rates, we need to specify a standard unit of time. It doesn't matter what we choose, but to make life easier, let's express rates as per minute. So if $q_{0,1} = 0.2$, then there is roughly 20% chance that the system currently in state 0 will be in state 1 after one minute.

Determining transition rates Let's consider a house of just one person. If the person is not currently in the bathroom, what is the transition rate of them going to the bathroom? A quick Google search suggests that people go to the toilet around 7 times a day. Taking into account time spent sleeping (approximated at 8 hours), that's a rate of 7 times per 16 hours, or $\frac{7}{(16 \times 60)} = 0.0073$ per minute. Denote this as λ.

Bathroom trips take a varying amount of time. For simplicity, let's assume an average trip takes 5 minutes. This implies a certain transition rate of leaving the bathroom. In our Markov chain model, the transition rate out of a state is equal to the inverse of the expected time spent in the state. If we let this transition rate be denoted by μ, then $\mu = \frac{1}{5} = 0.2$

Now that we have estimates for λ the rate of one person going to the bathroom, and μ, the rate of one person leaving the bathroom, we can easily determine the transition rate between states.
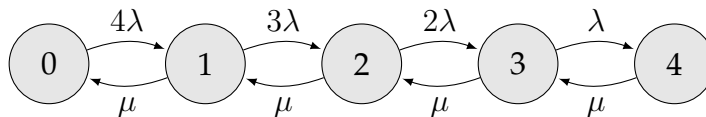

Figure 1: Diagram showing transition rates between the 5 states.

In state 0, there are 4 people that can potentially need to use the bathroom, each with a transition rate of λ, meaning $q_{0,1} = 4\lambda$. In state 1, there are only 3 people that could potentially transition, so $q_{1,2} = 3\lambda$

Using similar logic, we get the general expression:

$$q_{i,i+1} = (4 - i)\lambda \qquad \text{for } i < 4$$

What about the transition rates of leaving the bathroom? In any state apart from 0, one person is using the bathroom, and the rate of transition for that person is μ, leaving us with

$$q_{i,i-1} = \mu \qquad \text{for } i > 0$$

It is assumed that multiple people can't suddenly need to use the bathroom or leave the bathroom at the exact same time, so each state can only reach the state one above or below it, so:

$$q_{i,j} = 0 \qquad \text{for } |i - j| > 1$$

$q_{i,i}$ represents the rate at which the state stays in the same state. This is the same as saying that the state does not transition to any other state, so it is the negative of the sum of all transition rates out of the state:

$$q_{i,i} = -\sum_{j \neq i} q_{i,j}$$

Transition rate matrix Once these transition rates have been determined, they can be arranged in matrix form. Each column and row represent a state, and each element of the matrix is the transition rate of moving from the state represented by the row, to the state represented by the column.

$$Q = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{pmatrix} -4\lambda & 4\lambda & 0 & 0 & 0 \\ \mu & -(\mu + 3\lambda) & 3\lambda & 0 & 0 \\ 0 & \mu & -(\mu + 2\lambda) & 2\lambda & 0 \\ 0 & 0 & \mu & -(\mu + \lambda) & \lambda \\ 0 & 0 & 0 & \mu & -\mu \end{pmatrix}$$

State: 0, 1, 2, 3, 4

And after substituting $\mu = 0.2$ and $\lambda = \frac{7}{(16 \times 60)}$, we get

$$Q = \begin{pmatrix} -0.029 & 0.029 & 0 & 0 & 0 \\ 0.2 & -0.222 & 0.022 & 0 & 0 \\ 0 & 0.2 & -0.215 & 0.015 & 0 \\ 0 & 0 & 0.2 & -0.207 & 0.007 \\ 0 & 0 & 0 & 0.2 & -0.2 \end{pmatrix}$$

You can verify a few things about this transition matrix:

3

- The sum of each row equals 0. This is because the "rate" at which the state stays the same is exactly negative of the sum of the rates at which the state leaves.

- The only negative elements are on the diagonal. These represents the rate at which the state stays the same.

Long run proportions

Now that the system is fully described, it is time to think about our original problem. Living in a house of 4 people with 1 bathroom, what is the expected time one could expect to have to wait for the bathroom to be vacant each day? The states where someone is waiting to use the bathroom are states 2, 3 and 4, so let's start by calculating the proportion of time we expect this system to be in these states. To do this, we need to find the long run proportions of each state.

Let π_i represent the proportion of time the system is spend in state i if run forever. Since these are proportions, we have $\sum \pi_i = 1$

In the long run, the system will reach an equilibrium where everything balances perfectly. In this context, balance can be represented as:

$$\pi_i \times -q_{i,i} = \sum_{j \neq i} \pi_j \times q_{j,i}$$

On the left hand side, we have the rate of transition OUT of state i as time goes to infinity. On the right, we have the rate of transition INTO state i as time goes to infinity. Equating these imply we have reached the equilibrium.

Another way of writing this is

$$0 = \pi_i \times q_{i,i} + \sum_{i \neq j} \pi_j \times q_{j,i} \qquad \text{for all } i$$

Now let's define the vector $\boldsymbol{\pi} = (\pi_0, \cdots, \pi_4)$. Another way to write the above equations is as a homogenous system of linear equations:

$$\boldsymbol{\pi}Q = \mathbf{0}$$

Where we have the constraint $\sum \pi_i = 1$. Another way of writing this constraint is $\boldsymbol{\pi} \cdot \mathbf{1} = 1$ where $\mathbf{1}$ is a row vector of ones. Yet another way or writing this is

$$\boldsymbol{\pi}E = \mathbf{1}$$

Where E is a 5 by 5 matrix whose elements are all 1, $E = \begin{pmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{pmatrix}$

$\boldsymbol{\pi}$ can then be solved:

$$\boldsymbol{\pi}Q = \mathbf{0}$$

$$\pi E = \mathbf{1}$$
$$\pi(Q + E) = \mathbf{1}$$
$$\pi = \mathbf{1}(Q + E)^{-1}$$

Computer packages can be used to solve

$$(Q+E)^{-1} = \begin{pmatrix} 0.971 & 1.029 & 1 & 1 & 1 \\ 1.2 & 0.778 & 1.022 & 1 & 1 \\ 1 & 1.2 & 0.785 & 1.015 & 1 \\ 1 & 1 & 1.2 & 0.793 & 1.007 \\ 1 & 1 & 1 & 1.2 & 0.8 \end{pmatrix}^{-1}$$

To get us to our answer:

$$\pi = (8.599 \times 10^{-1}, 1.254 \times 10^{-1}, 1.372 \times 10^{-2}, 1.000 \times 10^{-3}, 3.646 \times 10^{-5})$$

Meaning that 86 % of the time, the bathrooms is empty, and 12.5 % of the time, there is one person using a bathroom and no one waiting. You can verify that these proportions add up to very close to 1 (error coming from rounding).

From this, we can sum $\pi_2 + \pi_3 + \pi_4$ to see that 1.48% of the time, there is at least one person waiting to use the bathroom. Out of our 16 waking hours, this corresponds to about 14 minutes a day where at least one person is waiting to use the bathroom.

To find out the amount of time per person, it might be tempting to simply divide that number by 4, but it's not quite so simple. To demonstrate why, consider two people waiting. You start waiting at 12:00, and manage to go at 12:15, your brother starts waiting as 12:05 and manages to go at 12:20. You've each waited 15 minutes, but the total amount of minutes any person was waiting in the house is only 20 minutes. Although you were both waiting from 12:05 to 12:15, this measure doesn't double up the wait time.

You must calculate the sum of everyone's wait time and divide it by the number of people. In the above example, we had 2 people waiting 15 minutes each, so the calculation would be $(15 + 15)/2$.

To find the sum of everyone's waiting time, notice that 1.372% of the time, there is one person waiting, 0.1% of the time, 2 people are waiting, and 0.003646% of the time, 3 people are waiting. In other words, to get the sum of everyone's waiting time, we do

$$(1.372\% \times 1 + 0.1\% \times 2 + 0.003646\% \times 3) \times 16 \times 60 \div 4 = 3.8$$

Leaving each housemate with an expected 3.8 minutes (3 minutes 48 seconds) of waiting time per day.

Waiting time under different scenarios

The beauty of having a mathematical model is that it can shed insight into the system under different conditions. The calculation above assumed 4 people and 1 bathroom, let's see what the waiting time would be with different numbers of people and bathrooms.

Expected minutes per day waiting to use the bathroom

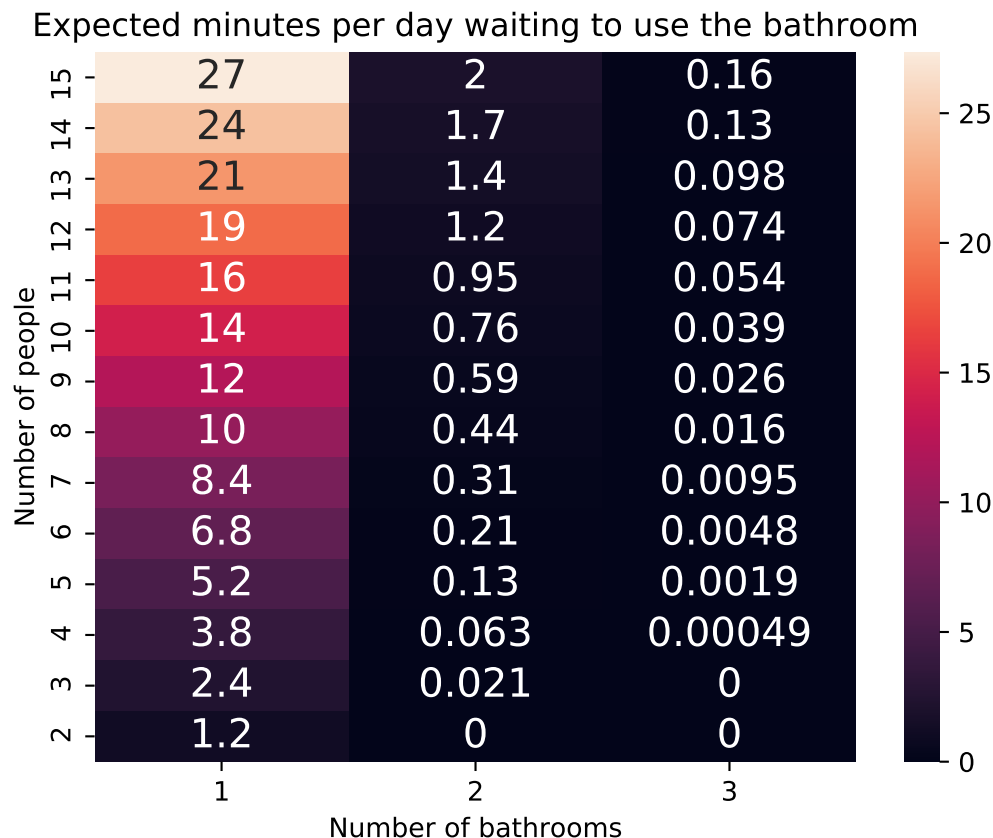| Number of people	1	2	3
15	27	2	0.16
14	24	1.7	0.13
13	21	1.4	0.098
12	19	1.2	0.074
11	16	0.95	0.054
10	14	0.76	0.039
9	12	0.59	0.026
8	10	0.44	0.016
7	8.4	0.31	0.0095
6	6.8	0.21	0.0048
5	5.2	0.13	0.0019
4	3.8	0.063	0.00049
3	2.4	0.021	0
2	1.2	0	0

Number of bathrooms

Figure 2: Comparing the average waiting time for each person under different scenarios

What's interesting to see is that an additional person doesn't add too much to the expected daily waiting time, and adding an extra bathroom dramatically reduces the expected waiting time. Under this model, 3 people sharing one bathroom spend more waiting time than 15 people with 2 bathrooms!

Limitations of the model

Now that there is a numeric answer to our question, it is important to look back on the process and acknowledge the limitations.

Memoryless   One of the main assumptions of Markov chains is the Markov property. This is also referred to as *memoryless* because it doesn't matter what's happened in the past, the behaviour of the system only depends on what the current state is.

We've assumed that the expected time spent in the bathroom is 5 minutes. So as soon as someone walks into the bathroom, it is expected they'll leave after 5 minutes. Under the Markov property, if 5 minutes passes and they're still in the bathroom, the expected time left is still 5 minutes.

Similarly, the rate at which someone goes to the bathroom stays the same regardless of if they've gone 10 times already or haven't gone at all.

While memoryless is not necessarily a realistic assumptions, the final implication is that bathroom times are *exponentially distributed* and the number of bathroom trips per day is *Poisson distributed* and this is quite a reasonable assumption.

Independence   Another assumption of the Markov chain model is independence. This assumption states that the each person acts independently of everyone else, and that the time spent in the bathroom is independent of everything. In real life, this assumption is likely not held for. Here are just a few examples of scenarios where independence isn't satisfied

- Most people want to use the bathroom at similar times during the day, such as in the morning after waking up and at night before bed, meaning the rate of using the bathroom is dependent on the time of day

- The expected time in the bathroom could be dependent on the time of day as most people shower in the morning or night.

- The expected time in the bathroom could be dependent on what people have already done in the bathroom today (e.g. already showered)

- After a night out, people may get home at the same time and all want to use the bathroom.

Code

All calculations were done in Python 3. Below is the code used to calculate the numbersin figure 2 and display them as the heatmap seen.

```python
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

def min_waiting(n, k, trips_per_day=7, avg_minutes_spent=5, waking_hours=16
                ):
    ''' n people in a house with k bathrooms.
    lambda_ times each person on average uses the bathroom per minute.
```

```python
    mu minutes spent on average in the bathroom per use.
    Returns the number of minutes each person can expect to wait per day
    '''

    # Rate of going into bathroom, underscore as lambda is Python keyword
                                        17/24 represents the hours we are
                                        awake
    lambda_ = trips_per_day / waking_hours / 60.0
    mu = 1/avg_minutes_spent # Rate of leaving

    # Transition matrix
    Q = np.zeros((n + 1, n + 1))
    for i in range(n + 1):
        if i < n:
            Q[i][i+1] = (n-i) * lambda_
        if i > 0:
            Q[i][i-1] = min(k, i) * mu

        Q[i][i] = - sum(Q[i, :])

    E = np.ones(shape=(n+1, n+1))
    pi = np.matmul(np.ones(shape=(1, n+1)), np.linalg.inv(Q + E))
    pi = pi.reshape(n+1)
    assert np.all(np.matmul(pi, Q) < 1e-10)

    ppl_minutes = 0
    for ppl_waiting, time in enumerate(pi[k:]):
        ppl_minutes += ppl_waiting * time
    prop_wait = ppl_minutes/n

    min_wait_per_day = prop_wait * waking_hours * 60
    return min_wait_per_day

########################################################################
# Simulations, plot heatmap
########################################################################
MAX_PEOPLE = 15
MAX_BATHROOM = 3

results = pd.DataFrame(0.0, index=range(MAX_PEOPLE, 1, -1),
  columns=range(1, MAX_BATHROOM + 1))
results.index.name = 'Num_people'

for num_people in range(2, MAX_PEOPLE + 1):
    for num_bathroom in range(1, min(num_people, MAX_BATHROOM + 1)):
        results[num_bathroom][num_people] = min_waiting(num_people,
                                        num_bathroom)

print('Results calculated')
sns.heatmap(results, annot=True, annot_kws={"size": 15})
plt.xlabel('Number of bathrooms')
plt.ylabel('Number of people')
plt.title('Expected minutes per day waiting to use the bathroom')
```

```
plt.show()
```