Topics in
Probability Theory and Stochastic Processes
Steven R. Dunbar

Waiting Time to Absorption

Rating
Mathematically Mature: may contain mathematics beyond calculus with proofs.
Section Starter Question

For a Markov chain with an absorbing state, describe the random variable for the time until the chain gets absorbed.

Key Concepts

1. Let \( \{X_n\} \) be a finite-state absorbing Markov chain with \( a \) absorbing states and \( t \) transient states. Let the \((a + t) \times (a + t)\) transition probability matrix be \( P \). Order the states so the absorbing states come first and non-absorbing, i.e. transient, states come last. Then the transition probability matrix has the block-matrix form

\[
P = \begin{pmatrix} I_a & 0 \\ A & T \end{pmatrix}
\]  

Here \( I_a \) is an \( a \times a \) identity matrix, \( A \) is the \( t \times a \) matrix of single-step transition probabilities from the \( t \) transient states to the \( a \) absorbing states, \( T \) is a \( t \times t \) submatrix of single-step transition probabilities among the transient states, and \( 0 \) is a \( a \times t \) matrix of 0s representing the single-step transition probabilities from absorbing states to transient states.

2. The matrix \( N = (I - T)^{-1} \) is the fundamental matrix for the absorbing Markov chain. The entries \( N_{ij} \) of this matrix have a probabilistic interpretation. The entries \( N_{ij} \) are the expected number of times that the chain started from state \( i \) will be in state \( j \) before ultimate absorption.

3. First-step analysis gives a compact expression in vector-matrix form for the waiting time \( w \) to absorption:

\[
(I - T)w = 1
\]

so \( w = (I - T)^{-1}1 \).
Vocabulary

1. Let \( \{X_n\} \) be a finite-state absorbing Markov chain with \( a \) absorbing states and \( t \) transient states. Let the \((a + t) \times (a + t)\) transition probability matrix be \( P \). Order the states so the absorbing states come first and non-absorbing, i.e. transient, states come last. Then \( P_{ij}^{(n)} \to 0 \) as \( n \to \infty \) for \( i \) and \( j \) in the transient states, while for \( i \) in the absorbing states, \( P_{ii} = 1 \). Define the random absorption time

\[
T = \min \{n \geq 0 : X_n \geq r\}.
\]

2. The absorption probability matrix \( B \) is the probability of starting at state \( i \) and ending at a given absorbing state \( j \).

3. For a Markov chain with \( a \) absorbing states and \( t \) transient states, if necessary, reorder the states so the absorbing states come first and non-absorbing, i.e. transient, states come last. Then the transition matrix has the canonical form:

\[
P = \begin{pmatrix}
I_a & 0 \\
A & T
\end{pmatrix}.
\]

The matrix \( N = (I - T)^{-1} \) is the fundamental matrix for the absorbing Markov chain.

4. The \( N \)th harmonic number \( H_N \) is

\[
H_N = 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{5} + \cdots + \frac{1}{N}.
\]
Mathematical Ideas

Theory

Let \( \{ X_n \} \) be a finite-state absorbing Markov chain with \( a \) absorbing states and \( t \) transient states. Let the \((a+t) \times (a+t)\) transition probability matrix be \( P \). Order the states so the absorbing states come first and non-absorbing, i.e. transient, states come last. The states \( a+1, 1, 2, \ldots, a+t \) are transient in that \( P_{ij}^{(n)} \to 0 \) as \( n \to \infty \) for \( a+1 \leq i, j \leq a+t \), while states \( 1, \ldots, a \) are absorbing, \( P_{ii} = 1 \) for \( 1 \leq i \leq a \). Then the transition probability matrix has the block-matrix form

\[
P = \begin{pmatrix}
I_a & 0 \\
A & T
\end{pmatrix}.
\]

(2)

Here \( I_a \) is an \( a \times a \) identity matrix, \( A \) is the \( t \times a \) matrix of single-step transition probabilities from the \( t \) transient states to the \( a \) absorbing states, \( T \) is a \( t \times t \) submatrix of single-step transition probabilities among the transient states, and 0 is a \( a \times t \) matrix of 0s representing the single-step transition probabilities from absorbing states to transient states.

Starting at one of the transient states \( i \) where \( 1 \leq i \leq t \) such a process will remain in the transient states for some duration. Ultimately, the process gets trapped in one of the absorbing states \( i = 1, \ldots, a \). Before the Markov process transitions to one of the absorbing states, the number of times it visits a transient state is a random variable. Let \( Y_{ij} \) denote the number of visits the system makes to transient state \( j \) before reaching an absorbing state, given the system started in transient state \( i \). Thus, \( Y_{ij} \) is a discrete random variable that can take on any nonnegative integer value. The random variables \( Y_{ij} \) are the fundamental random variables of interest here. These fundamental random variables are the building blocks for constructing and investigating other random variables. The mean, variance and covariances of the \( Y_{ij} \) are the first statistics to investigate. Of special interest is the mean time until absorption. Define the random absorption time

\[
w_i = \min \{ n \geq 0 : X_n \leq a \mid X_0 = i \}.
\]

Notice that \( w_i = \sum_{k=a+1}^{a+t} Y_{ik} \), the total number of visits the process makes among the transient states. The expected value of this random time \( \mathbb{E}[w_i] \), \( i = a + 1, \ldots, a+t \) is a first measure of the random variable \( w_i \). Also of interest is the probability distribution of the states into which absorption takes
place. Using the fundamental random variables, it is possible to compute this probability too.

**Indicator Bernoulli Random Variables**

Let the indicator random variables be

\[ U_{ij}^{(m)} = \begin{cases} 
1 & \text{if the Markov chain is in transient state } j \\
0 & \text{if the Markov chain is not in transient state } j 
\end{cases} \]

after \( m \) steps given that it starts in transient state \( i \) for \( m = 0, 1, 2, \ldots \). The case \( m = 0 \) simply indicates where the system starts:

\[ U_{ij}^{(0)} = \begin{cases} 
1 & \text{if the Markov chain starts in transient state } i \\
0 & \text{if the Markov chain does not start in transient state } i.
\end{cases} \]

Using the usual notation, \( u_{ij}^{(0)} \) is the Kronecker delta function, \( \delta_{ij} \). The indicator random variables connect to the fundamental random variables \( Y_{ij} \) through the sum

\[ Y_{ij} = \sum_{m=0}^{\infty} U_{ij}^{(m)} . \]

**Expected number of visits between states**

The expected number of visits to transient state \( j \) given that the Markov chain starts in transient state \( i \) in terms of the indicator random variable is

\[ \mathbb{E}[Y_{ij}] = \mathbb{E} \left[ \sum_{m=0}^{\infty} U_{ij}^{(m)} \right] = \sum_{m=0}^{\infty} \mathbb{E} \left[ U_{ij}^{(m)} \right] . \]

Now use mathematical induction to show

\[ P^m = \begin{pmatrix} I_a & 0 \\ A & T \end{pmatrix}^m \begin{pmatrix} I_a & 0 \\ A & T \end{pmatrix} = \begin{pmatrix} I_a & 0 \\ (I_t + T + T^2 + \cdots + T^{m-1})A & T^m \end{pmatrix} . \]

The elements \( p_{ij}^{(m)} \) of \( P^m \) are the \( m \)-step transition probabilities between all states. Since \( \mathbb{E} \left[ U_{ij}^{(m)} \right] = p_{ij}^{(m)} \), so

\[ \mathbb{E}[Y_{ij}] = \sum_{m=0}^{\infty} \mathbb{E} \left[ U_{ij}^{(m)} \right] = \sum_{m=0}^{\infty} p_{ij}^{(m)} . \]
When \(i\) and \(j\) are transient states, then we only need to consider the entries in the transient corner matrix \(T^m\), so
\[
\mathbb{E}[Y_{ij}] = \sum_{m=0}^{\infty} (T^m)_{ij} = \left( \sum_{m=0}^{\infty} T^m \right)_{ij}.
\]

The basic theory of finite absorbing Markov chains ensures that the Euclidean norm of \(T\) is less than 1, \(\|T\| < 1\). Therefore \(\sum_{m=0}^{\infty} T^m\) converges. Furthermore, it converges to the fundamental matrix \(N = (I-T)^{-1}\). Thus \(\mathbb{E}[Y_{ij}] = N_{ij}\).

**Waiting time to absorption**

**Theorem 1.** The entries \(N_{ij}\) of the fundamental matrix are the expected number of times that the chain started from state \(i\) will be in state \(j\) before ultimate absorption and the vector of expected waiting times to absorption is \(\mathbb{E}[w] = N = (I-T)^{-1}1\).

**Proof.** The sum over all states \(j\) of the number of times that the chain started from state \(i\) will be in state \(j\) before ultimate absorption is the waiting time to absorption. \(\square\)

**Covariances of numbers of visits between states**

Now we want to use the indicator Bernoulli random variables to derive \(\text{Cov}[Y_{ij}, Y_{ik}]\) where \(i, j, k\) are transient states. Recall that
\[
\text{Cov}[Y_{ij}, Y_{ik}] = \mathbb{E}[Y_{ij} \cdot Y_{ik}] - \mathbb{E}[Y_{ij}] \cdot \mathbb{E}[Y_{ik}]
\]
and since we already know \(\mathbb{E}[Y_{ij}]\) and \(\mathbb{E}[Y_{ik}]\), all that is necessary is \(\mathbb{E}[Y_{ij} \cdot Y_{ik}]\). Start with
\[
Y_{ij}Y_{ik} = \left( \sum_{x=0}^{\infty} U_{ij}^{(x)} \right) \left( \sum_{y=0}^{\infty} U_{ik}^{(y)} \right) = \sum_{x=0}^{\infty} \sum_{y=0}^{\infty} U_{ij}^{(x)} U_{ik}^{(y)}
\]
so that
\[
\mathbb{E}[Y_{ij}Y_{ik}] = \mathbb{E} \left[ \sum_{x=0}^{\infty} \sum_{y=0}^{\infty} U_{ij}^{(x)} U_{ik}^{(y)} \right] = \sum_{x=0}^{\infty} \sum_{y=0}^{\infty} \mathbb{E} \left[ U_{ij}^{(x)} U_{ik}^{(y)} \right].
\]
Rearrange the double sum to

\[
\sum_{x=0}^{\infty} \sum_{y=x+1}^{\infty} E \left[ U_{ij}^{(x)} U_{ik}^{(y)} \right] + \sum_{x=0}^{\infty} \sum_{y=x}^{\infty} E \left[ U_{ij}^{(x)} U_{ik}^{(x)} \right] + \sum_{y=0}^{\infty} \sum_{x=y+1}^{\infty} E \left[ U_{ik}^{(y)} U_{ij}^{(x)} \right].
\]

Note that the first term sums over the lattice points above the line \( y = x \), the second term sums over the lattice points along the line \( y = x \), the third term sums over lattice points below the line \( y = x \). The first and third terms are symmetric, so we only evaluate the first term.

The expression \( E \left[ U_{ij}^{(x)} U_{ik}^{(y)} \right] \) is the probability that the system is in transient state \( i \) after exactly \( x \) steps from the start in state \( i \) and the system is in transient state \( k \) after exactly \( y \) steps from the start in state \( i \). Recall that \( x < y \). Using the Markov chain property, this is

\[
E \left[ U_{ij}^{(x)} U_{ik}^{(y)} \right] = p_{ij}^{(x)} p_{jk}^{(y-x)}.
\]

Therefore, the first term is

\[
\sum_{x=0}^{\infty} \sum_{y=x+1}^{\infty} E \left[ U_{ij}^{(x)} U_{ik}^{(y)} \right] = \sum_{x=0}^{\infty} \sum_{y=x+1}^{\infty} p_{ij}^{(x)} p_{jk}^{(y-x)}
\]

\[
= \sum_{x=0}^{\infty} \sum_{y=x+1}^{\infty} p_{ij}^{(x)} p_{jk}^{(y-x)}
\]

\[
= \sum_{x=0}^{\infty} \sum_{z=1}^{\infty} p_{ij}^{(x)} p_{jk}^{(z)}
\]

\[
= \left( \sum_{x=0}^{\infty} p_{ij}^{(x)} \right) \left( \sum_{z=0}^{\infty} p_{jk}^{(z)} \right)
\]

\[
= \left( \sum_{x=0}^{\infty} (T^x)_{ij} \right) \left( \sum_{z=0}^{\infty} (T^z)_{jk} - \delta_{jk} \right)
\]

\[
= \left( \sum_{x=0}^{\infty} (T^x)_{ij} \right) \left( \sum_{z=0}^{\infty} T^z \right)_{jk} - \delta_{jk}
\]

\[
= (I_t - T)^{-1}_{ij} (I_t - T)^{-1}_{jk} - \delta_{jk}
\]

\[
= N_{ij} (N_{jk} - \delta_{jk}).
\]
The third term is the first term with \( j \) and \( k \) interchanged, so
\[
\sum_{y=0}^{\infty} \sum_{x=y+1}^{\infty} \mathbb{E} \left[ U_{ik}^{(y)} U_{ij}^{(x)} \right] = N_{ik} (N_{kj} - \delta_{jk}).
\]

Finally, the second term is \( \sum_{x=0}^{\infty} \mathbb{E} \left[ U_{ij}^{(x)} U_{ik}^{(x)} \right] \) where each summand is the probability that the Markov chain is in transient state \( j \) after exactly \( x \) steps after starting in transient state \( i \) and simultaneously in state \( k \) after exactly \( x \) steps starting from transient state \( i \). This is only possible if \( j = k \) hence
\[
\mathbb{E} \left[ U_{ij}^{(x)} U_{ik}^{(x)} \right] = p_{ij}^m \delta_{jk}. \quad \text{Thus for the second term}
\]
\[
\sum_{x=0}^{\infty} \mathbb{E} \left[ U_{ij}^{(x)} U_{ik}^{(x)} \right] = \sum_{x=0}^{\infty} p_{ij}^m \delta_{jk} = \left( \sum_{x=0}^{\infty} (T^x)_{ij} \right) \delta_{jk} = \left( \sum_{x=0}^{\infty} T^x \right)_{ij} \delta_{jk} = N_{ij} \delta_{kj}.
\]

Putting all terms together
\[
\mathbb{E} \left[ Y_{ij} Y_{ik} \right] = N_{ij} (N_{kj} - \delta_{jk}) + N_{ij} \delta_{jk} + N_{ik} (N_{kj} - \delta_{jk}) = N_{ij} N_{jk} + N_{ik} N_{kj} - N_{ik} \delta_{kj}.
\]

Then
\[
\text{Cov} \left[ Y_{ij}, Y_{jk} \right] = \mathbb{E} \left[ Y_{ij} Y_{ik} \right] - \mathbb{E} \left[ Y_{ij} \right] \mathbb{E} \left[ Y_{ik} \right] = N_{ij} N_{jk} + N_{ik} N_{kj} - N_{ij} N_{ik} - N_{ik} \delta_{kj}.
\]

In particular, letting \( j = k \) we get the variance of the number of visits to state \( j \) starting from \( i \):
\[
\text{Var} \left[ Y_{ij} \right] = 2 N_{ij} N_{jj} - N_{ij}^2 - N_{ij}.
\]

Following the notation of Kemeny and Snell, if we set \( \text{diag}(N) \) to be the diagonal matrix setting all off-diagonal elements of \( N \) to 0, and if we set \( N_{sq} \) to be the matrix resulting from squaring each entry, then
\[
\text{Var} \left[ X \right] = N (2 \text{diag}(N) - I) - N_{sq}.
\]

For covariances, for a fixed \( i \), we want the \( t \times t \) matrix of entries \( \text{Cov} \left[ Y_{ij}, Y_{ik} \right] \). From the formula in terms of the fundamental matrix \( N \), the covariance is the product of row vectors of \( N \) with \( N \). Letting \( N_{(i)} \) denote the \( 1 \times t \) row-vector from row \( i \) and setting \( \text{diag} \ N_{(i)} \) to be the diagonal matrix with this vector along the diagonal this becomes
\[
\text{diag} \ N_{(i)} N + N \text{diag} \ N_{(i)} - N_{(i)}^T N_{(i)} - \text{diag} \ N_{(i)}.
\]

Note the outer product \( N_{(i)}^T N_{(i)} \).
First-step analysis

First-step analysis tells us that the absorption time from state \( i \) is the first step to another transient state plus a weighted average, according to the transition probabilities over the transient states, of the absorption times from the other transient states. In symbols, first-step analysis says

\[
w_i = 1 + \sum_{j=a+1}^{a+t} P_{ij} w_j.
\]

As before, for a Markov chain with \( a \) absorbing states and \( t \) transient states, reorder the states so the absorbing states come first and non-absorbing, i.e. transient, states come last. Then the transition matrix has the canonical form:

\[
P = \begin{pmatrix} I_a & 0 \\ A & T \end{pmatrix}.
\]

Here \( I_a \) is an \( a \times a \) identity matrix, \( A \) is the \( t \times a \) matrix of single-step transition probabilities from the \( t \) transient states to the \( a \) absorbing states, \( T \) is a \( t \times t \) submatrix of single-step transition probabilities among the transient states, and 0 is a \( a \times t \) matrix of 0s representing the single-step transition probabilities from absorbing states to transient states.

Expressing the first-step analysis compactly in vector-matrix form as

\[
w = 1 + Tw
\]

or

\[(I - T)w = 1\]

then \( w = (I - T)^{-1} \mathbf{1} \). The matrix \( N = (I - T)^{-1} \) is the fundamental matrix for the absorbing Markov chain. The entries \( N_{ij} \) of this matrix have a probabilistic interpretation. The entries \( N_{ij} \) are the expected number of times that the chain started from state \( i \) will be in state \( j \) before ultimate absorption.

The \( t \times a \) matrix of absorption probabilities \( B \) has as entries the probability of starting at state \( i \) and ending up at a given absorbing state \( j \). The absorption probabilities come from \( N \) by the matrix product \( B = NA = (I - T)^{-1}A \). Roughly, add up all the probabilities of going to state \( j \), weighted by the number of times we expect to be in the transient states. More precisely, we have the following theorem.
Theorem 2. Let $b_{ij}$ be the probability of the Markov process starting in transient state $i$ and ending in absorbing state $j$, then

$$B = (b_{ij}) = NA.$$ 

Proof. The proof is by first-step analysis. Starting in state $i$, the process may be captured in $j$ in one or more steps. The probability of capture in a single step is $p_{ij}$. If this does not happen, the process may move to another absorbing state, in which case it is impossible to reach $j$, or to a transient state $k$. In the latter case, the probability of being captured in $j$ is $b_{kj}$. Hence

$$b_{ij} = p_{ij} + \sum_{k=a+1}^{a+t} p_{ik} b_{kj}$$

or in matrix form $B = A + TB$. Thus, $B = (I - T)^{-1} A = NA$. 

Examples

Example. Consider a random walk of a particle which moves in a straight line in unit steps. Each step is 1 unit to the right with probability $p$ and to the left with probability $q$. It moves until it reaches one of two extreme points which are boundary points. Assume that if the process reaches the boundary points, it remains there from that time on. Figure has 9 states numbered from $-4$ to $4$. The absorbing boundary states are $-4$ and $4$. 

Figure 1: Image of a possible random walk in phase line after an odd number of steps.
The full transition probability matrix is

\[
P = \begin{pmatrix}
-4 & -3 & -2 & -1 & 0 & 1 & 2 & 3 & 4 \\
-4 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
-3 & q & 0 & p & 0 & 0 & 0 & 0 & 0 \\
-2 & 0 & q & 0 & p & 0 & 0 & 0 & 0 \\
-1 & 0 & 0 & q & 0 & p & 0 & 0 & 0 \\
0 & 0 & 0 & q & 0 & p & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & q & 0 & p & 0 & 0 \\
2 & 0 & 0 & 0 & 0 & q & 0 & p & 0 \\
3 & 0 & 0 & 0 & 0 & 0 & q & 0 & p \\
4 & 0 & 0 & 0 & 0 & 0 & 0 & q & 0
\end{pmatrix}
\]

Reorder the states as \(-4, 4, -3, -2, -1, 0, 1, 2, 3\) to bring the transition probability matrix to standard form

\[
P_1 = \begin{pmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
q & 0 & 0 & p & 0 & 0 & 0 & 0 & 0 \\
0 & q & 0 & p & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & q & 0 & p & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & q & 0 & p & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & q & 0 & p & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & q & 0 & p & 0 \\
0 & p & 0 & 0 & 0 & 0 & 0 & q & 0
\end{pmatrix}
\]

so

\[
T = \begin{pmatrix}
0 & p & 0 & 0 & 0 & 0 & 0 \\
q & 0 & p & 0 & 0 & 0 & 0 \\
0 & q & 0 & p & 0 & 0 & 0 \\
0 & 0 & q & 0 & p & 0 & 0 \\
0 & 0 & 0 & q & 0 & p & 0 \\
0 & 0 & 0 & 0 & q & 0 & p \\
0 & 0 & 0 & 0 & 0 & q & 0
\end{pmatrix}
\]

A computer algebra system such as Maxima for example can compute the fundamental matrix. Two representative examples will suffice to show the possibilities.
Then the waiting times to absorption are 7, 12, 15, 16, 15, 12, 7. For the variances, consider only the central state, originally labeled as 0 and after re-ordering is the sixth state in the middle of the transient states.

The variance is $1^T V 1 = 160$, the standard deviation is $4\sqrt{10}$.

The probabilities of absorption are

$$N(1/2)A = \begin{pmatrix}
7 & \frac{5}{4} & \frac{5}{2} & \frac{3}{2} & 1 & \frac{3}{2} & \frac{1}{2} & \frac{1}{2} \\
\frac{3}{4} & \frac{3}{2} & \frac{5}{2} & \frac{5}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\
1 & 2 & 3 & 4 & 3 & 2 & 1 \end{pmatrix}$$

which are symmetric as expected.

If $p = 2/3$ so the probability of moving to the right is twice the probability
of moving to the left, then

\[
N(2/3) = \begin{pmatrix}
127 & 126 & 124 & 24 & 112 & 96 & 64 \\
85 & 85 & 85 & 17 & 17 & 17 & 17 \\
31 & 31 & 31 & 11 & 11 & 11 & 11 \\
7 & 7 & 7 & 3 & 3 & 3 & 3 \\
3 & 3 & 3 & 1 & 1 & 1 & 1 \\
85 & 85 & 85 & 85 & 85 & 85 & 85
\end{pmatrix}
\]

and the waiting times to absorption are 9.047058823529412, 12.07058823529412, 12.08235294117647, 10.58823529411765, 8.341176470588236, 5.71764705882353, 2.905882352941176. For the variances, consider only the central state, originally labeled as 0 and after reordering is the sixth state in the middle of the transient states.

\[
\text{Cov} \begin{pmatrix} Y_{0j}, Y_{0k} \end{pmatrix} = \begin{pmatrix}
127 & 126 & 124 & 24 & 112 & 96 & 64 \\
85 & 85 & 85 & 17 & 17 & 17 & 17 \\
31 & 31 & 31 & 11 & 11 & 11 & 11 \\
7 & 7 & 7 & 3 & 3 & 3 & 3 \\
3 & 3 & 3 & 1 & 1 & 1 & 1 \\
85 & 85 & 85 & 85 & 85 & 85 & 85
\end{pmatrix}
\]

The variance is \(1^T V 1 = 21708/289 \approx 75.11418\), the standard deviation is 8.66684, less than the symmetric case where \(p = 1/2 = q\), as expected.

The probabilities of absorption are

\[
N(2/3)A = \begin{pmatrix}
127 & 128 \\
85 & 85 \\
255 & 255 \\
1 & 1 \\
255 & 255
\end{pmatrix}
\]

The absorption probabilities are strongly biased to the right, as expected.

Example. A law firm employs three types of lawyers: junior lawyers, senior lawyers, and partners. During a given year, there is probability 0.15 of
promoting a junior lawyer to senior lawyer and probability 0.5 that the junior lawyer will leave the firm. There is probability 0.20 of promoting the senior lawyer to partner and probability 0.10 that the senior lawyer will leave the firm. Finally, there is probability 0.05 that a partner will leave the firm, see Table 1. The firm never demotes a lawyer or a partner. What is the average number of years that a newly hired junior lawyer stays with the firm?

Table 1: The transition probabilities

<table>
<thead>
<tr>
<th></th>
<th>leave firm</th>
<th>junior lawyer</th>
<th>senior lawyer</th>
<th>partner</th>
</tr>
</thead>
<tbody>
<tr>
<td>leave firm</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>junior lawyer</td>
<td>0.05</td>
<td>0.80</td>
<td>0.15</td>
<td>0</td>
</tr>
<tr>
<td>senior lawyer</td>
<td>0.10</td>
<td>0</td>
<td>0.70</td>
<td>0.20</td>
</tr>
<tr>
<td>partner</td>
<td>0.05</td>
<td>0</td>
<td>0</td>
<td>0.95</td>
</tr>
</tbody>
</table>

This leads to the single-step transition matrix $P$ partitioned into one absorbing state for “leave firm” and three transient states: junior lawyer, senior lawyer, partner.

$$\begin{pmatrix}
1 & 0 & 0 & 0 \\
0.05 & 0.80 & 0.15 & 0 \\
0.10 & 0 & 0.70 & 0.20 \\
0.05 & 0 & 0 & 0.95
\end{pmatrix}$$

The transient transition matrix is

$$\begin{pmatrix}
T = 0.80 & 0.15 & 0 \\
0 & 0.70 & 0.20 \\
0 & 0 & 0.95
\end{pmatrix}$$

The corresponding fundamental matrix is

$$N = (I_3 - T)^{-1} = \begin{pmatrix} 5 & 5/2 & 10 \\ 0 & 10/3 & 40/3 \\ 0 & 0 & 20 \end{pmatrix}$$

The waiting times to absorption are then

$$(I - T)^{-1}1 = (35/2, 50/3, 20)^T.$$
Now compute the covariances for the first transient state, junior lawyer. Using
\[
\text{diag } N(1)N + N \text{ diag } N(1) - N^T(1)N(1) - \text{diag}(N(1))
\]
we get
\[
\text{Cov } [Y_{1j}, Y_{1k}] = \begin{pmatrix} 20 & 0 & 0 \\ 0 & 95/12 & 25/3 \\ 0 & 25/3 & 290 \end{pmatrix} = \begin{pmatrix} 20 & 0 & 0 \\ 0 & 7.91667 & 8.33333 \\ 0 & 8.33333 & 290 \end{pmatrix}.
\]

Since \( w_i = \sum_{j=a+1}^t Y_{ij} \), using that \( \text{Var } [w_i] = \sum_{j=a+1}^t \sum_{k=a+1}^t \text{Cov } [Y_{1j}, Y_{1k}] \), the variance is the sum of the entries in the covariance matrix. Hence \( \text{Var } [w_1] = 334.48 \) (in units of \( \text{textyear}^2 \)). The standard deviation is 18.292 years. As a point of special interest, note that the standard deviation is larger than the mean so that knowing the mean is not sufficient to predict how long a junior lawyer will be with the firm.

*Example.* It’s your 30th birthday, and your friends bought you a cake with 30 candles on it. You make a wish and try to blow them out. Every time you blow, you blow out a random number of candles between one and the number that remain, including one and that other number. How many times on average do you blow before you extinguish all the candles?

Let the states be the number of candles blown out, so the \( N = 31 \) states are 0, 1, 2, 3, ..., 30. Instead of solving this at once, try a smaller problem first, with the number of candles \( N = 5 \). Then the transition probability matrix in canonical form is
\[
P = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} \\ 0 & 0 & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} \\ 0 & 0 & 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ 0 & 0 & 0 & 0 & \frac{1}{2} \end{pmatrix}
\]

Then the first-step equations for the waiting time, that is the number of
attempts needed to blow out the candles are

\[
\begin{align*}
w_0 &= 1 + \frac{1}{5}w_1 + \frac{1}{5}w_2 + \frac{1}{5}w_3 + \frac{1}{5}w_4 \\
w_1 &= 1 + \frac{1}{4}w_2 + \frac{1}{4}w_3 + \frac{1}{4}w_4 \\
w_2 &= 1 + \frac{1}{3}w_3 + \frac{1}{3}w_4 \\
w_3 &= 1 + \frac{1}{2}w_4 \\
w_4 &= 1.
\end{align*}
\]

Solving this recursively, \( w_4 = 1, \ w_3 = 1 + \frac{1}{2}, \ w_2 = 1 + \frac{1}{3}(1 + \frac{1}{2}) + \frac{1}{3}, \) so \( w_2 = 1 + \frac{1}{2} + \frac{1}{3}. \) Continuing to solve recursively

\[
\begin{align*}
w_1 &= 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} \\
w_0 &= 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{5}.
\end{align*}
\]

The inductive pattern is clear. The waiting time with \( N \) candles is the \( N \)th harmonic number \( H_N \)

\[
w_0 = H_N = 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{5} + \cdots + \frac{1}{N}.
\]

For the original problem with 30 candles the transient states are 1, 2, 3, \ldots, 30 and the absorbing state is 0. Then the transition probability matrix as in \([2]\) is

\[
P = \begin{pmatrix}
0 & \frac{1}{30} & \frac{1}{30} & \cdots & \frac{1}{30} \\
0 & 0 & \frac{1}{29} & \cdots & \frac{1}{29} \\
0 & 0 & 0 & \cdots & \frac{1}{28} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \cdots & \frac{1}{2} \\
0 & 0 & 0 & \cdots & 0 & 1
\end{pmatrix}
\]

The quantity of interest is the expected waiting time until absorption into the state 0. Expressing the first-step analysis compactly in vector-matrix form as

\[(I - T)v = 1\]
Substituting in the values from the transition matrix and solving with Octave,
this is \( w_0 = H_{30} \approx 3.9950 \).

On May 9, I get that the variance of the waiting time from the first transient state (all candles?) is 2.3828

**Example.** An urn contains two unpainted balls. At a sequence of times, choose a ball at random, then paint it either red or black, and put it back. For an unpainted ball, choose a color at random. For a painted ball, change its color. Form a Markov chain by taking as a state the triple \((x, y, z)\) where \(x\) is the number of unpainted balls, \(y\) the number of red balls, and \(z\) the number of black balls. The transition matrix is then

\[
\begin{pmatrix}
(0, 1, 1) & (0, 2, 0) & (0, 0, 2) & (2, 0, 0) & (1, 1, 0) & (1, 0, 1) \\
(0, 1, 1) & 0 & 1/2 & 1/2 & 0 & 0 & 0 \\
(0, 2, 0) & 1 & 0 & 0 & 0 & 0 & 0 \\
(0, 0, 2) & 1 & 0 & 0 & 0 & 0 & 0 \\
(2, 0, 0) & 0 & 0 & 0 & 1/2 & 1/2 \\
(1, 1, 0) & 1/4 & 1/4 & 0 & 0 & 0 & 1/2 \\
(1, 0, 1) & 1/4 & 0 & 1/4 & 0 & 1/2 & 0
\end{pmatrix}
\]

In this case, there is no absorbing state, that is, a state which once entered remains the same thereafter. However, the first three states together are an ergodic set, that is, once the process enters that set, it continues in that set. So lump those states together as a single absorbing state, and the transient states are \((2, 0, 0)\), \((1, 1, 0)\), and \((1, 0, 1)\). Then

\[
T = \begin{pmatrix}
0 & 1/2 & 1/2 \\
0 & 0 & 1/2 \\
0 & 1/2 & 0
\end{pmatrix}
\]

and the fundamental matrix is

\[
N = \begin{pmatrix}
1 & 1 & 1 \\
0 & \frac{4}{3} & \frac{2}{3} \\
0 & \frac{1}{3} & \frac{1}{3}
\end{pmatrix}
\]

With this, we can compute

\[
\mathbb{E}[w] = N1 = (3, 2, 2)^T
\]

and

\[
\text{Var}[w] = (2N - I) - (\mathbb{E}[1])_{sq} = (2, 2, 2)^T
\]
Finally,

$$\text{Var} [X] = N (2 \text{diag}(N) - I) - N_{sq} = \begin{pmatrix} 0 & 1/2 & 2/3 \\ 0 & 4/9 & 2/3 \\ 0 & 2/3 & 4/9 \end{pmatrix}$$

Since the process must immediately leave state $(2, 0, 0)$ and cannot go back, there is 0 variance for the number of times in this state.

*Example.* The following is a larger example of the painting the balls puzzle.

You play a game with four balls in a box: One ball is red, one is blue, one is green and one is yellow. You draw a ball out of the box at random and note its color. Without replacing the first ball, you draw a second ball and then paint it to match the color of the first. Replace both balls, and repeat the process. The game ends when all four balls have become the same color. What is the expected number of turns to finish the game? Extra credit: What if there are more balls and more colors?

Take the states as the number of balls of each different color without regard for the colors themselves. For example, partition 4 balls in 5 different ways:

$$1 + 1 + 1 + 1, \quad 2 + 1 + 1, \quad 2 + 2, \quad 3 + 1, \quad 4.$$  

The partition $2 + 1 + 1$, for example, consists of cases where two of the balls are the same color and the other two balls are two other colors. For example, the cases “red & red & green & blue” and “blue & blue & yellow & red’ correspond to the partition $2 + 1 + 1$. By using these five partitions as states in a Markov Chain, we can compute the transition probabilities to go from one state to the next. For example, the probability of transitioning from $2 + 1 + 1$ to $3 + 1$ is $\frac{1}{3}$ because in order for this transition to occur, we must first choose one of the two identically colored balls with probability $\frac{2}{4}$, then we must choose one of the other two balls out of the remaining three with probability $\frac{2}{3}$. The joint probability is $\frac{2}{4} \cdot \frac{2}{3} = \frac{1}{3}$. As another example, the probability of transitioning from $2 + 2$ to $3 + 1$ is the probability of first picking a ball of either color, leaving 1 ball of that color and 2 balls of the other color, then from those 3 balls picking the second color with probability $\frac{2}{3}$. Calculate all remaining transition probabilities in the same way.
The absorbing state is 4. Rearranging into the standard block-matrix form

\[
P = \begin{pmatrix}
1 + 1 + 1 + 1 & 2 + 1 + 1 & 3 + 1 & 2 + 2 & 4 \\
1 + 1 + 1 + 1 & 0 & 1 & 0 & 0 & 0 \\
2 + 1 + 1 & 0 & 1/2 & 1/3 & 1/6 & 0 \\
3 + 1 & 0 & 1/4 & 1/2 & 0 & 1/4 \\
2 + 2 & 0 & 0 & 2/3 & 1/3 & 0 \\
4 & 0 & 0 & 0 & 0 & 1
\end{pmatrix}
\]

Then \( N = (I - T)^{-1} \) is

\[
\begin{pmatrix}
4 & 1 + 1 + 1 + 1 & 2 + 1 + 1 & 3 + 1 & 2 + 2 \\
4 & 1 & 0 & 0 & 0 & 0 \\
1 + 1 + 1 + 1 & 0 & 0 & 1 & 0 & 0 \\
2 + 1 + 1 & 0 & 0 & 1/2 & 1/3 & 1/6 \\
3 + 1 & 1/4 & 0 & 1/4 & 1/2 & 0 \\
2 + 2 & 0 & 0 & 0 & 2/3 & 1/3
\end{pmatrix}
\]

and the waiting time from the initial state is \( 1 + 4 + 4 + 1 = 10 \) (Lessard gets 9 but maybe he doesn’t count initial state?). From the state \( 1 + 1 + 1 + 1 \) (labeled as 2 in the standard format), the covariance matrix is

\[
\begin{pmatrix}
1 & 4 & 4 & 1 \\
0 & 4 & 4 & 1 \\
0 & 2 & 4 & \frac{1}{2} \\
0 & 2 & 4 & 2
\end{pmatrix}
\]

Then the variance of the number of visits is \( \frac{175}{2} \) and the standard deviation of the number of visits is \( 5\sqrt{\frac{14}{2}} \approx 9.3541 \).
Sources
This section is adapted from Carchidi and Higgins, \[1\]. Other ideas are adapted from *Finite Markov Chains* by Kemeny and Snell, *An Introduction to Stochastic Modeling* by Taylor and Karlin and *Random Walks and Electrical Networks* by Doyle and Snell. The birthday candle example is adapted from the January 13, 2017 “Riddler” problem from the webpage fivethirtyeight.com.

Colorful balls from http://www.laurentlessard.com/bookproofs/colorful-balls-puzzle/

Algorithms, Scripts, Simulations

Algorithm

Scripts

Commands are

\[
T = \text{diag}(1 ./ [30:-1:1]) * \text{triu}(\text{ones}(30, 30)) - \text{eye}(30);
\]

\[
\text{wait} = \text{inverse}(\text{eye}(30) - T) * \text{ones}(30, 1);
\]

\[
\text{HN} = \sum(1 ./ [30:-1:1])
\]

Problems to Work for Understanding

1.
References


Outside Readings and Links:


2.

3.
4.

I check all the information on each page for correctness and typographical errors. Nevertheless, some errors may occur and I would be grateful if you would alert me to such errors. I make every reasonable effort to present current and accurate information for public use, however I do not guarantee the accuracy or timeliness of information on this website. Your use of the information from this website is strictly voluntary and at your risk.

I have checked the links to external sites for usefulness. Links to external websites are provided as a convenience. I do not endorse, control, monitor, or guarantee the information contained in any external website. I don’t guarantee that the links are active at all times. Use the links here with the same caution as you would all information on the Internet. This website reflects the thoughts, interests and opinions of its author. They do not explicitly represent official positions or policies of my employer.

Information on this website is subject to change without notice.

Steve Dunbar’s Home Page, http://www.math.unl.edu/~sdunbar1
Email to Steve Dunbar, sdunbar1 at unl dot edu
Last modified: Processed from \LaTeX{} source on June 10, 2017