STAT 202C - HW 3

John Rapp Farnes / 405461225

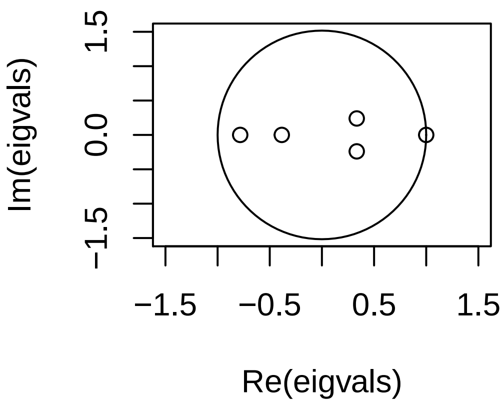
5/12/2020

1.

1.1.

a)

```
library(plotrix)
library(matrixcalc)

K <- t(matrix(c(
  .3,.6,.1,.0,.0,
  .2,.0,.7,.0,.1,
  .0,.5,.0,.5,.0,
  .0,.0,.4,.1,.5,
  .4,.1,.0,.4,.1
), nrow=5, ncol=5))

eig <- eigen(K)
eigvals <- eig$values

plot(eigvals, ylim = c(-1.5, 1.5), xlim = c(-1.5, 1.5))

draw.circle(0, 0, 1)
```



b)

K is primitive, and hence the invariant probability can be calculated by Perron-Frobenius Theorem as the left-eigenvector corresponding to the largest eigenvalue.

```r
# Left eigenvector corresponding to largest eigenvalue is pi
p <- Re(eigen(t(K))$vectors[,1])

# Make stochastic
p <- p / sum(p)

# Multiply with K^1000 to improve estimate
pi <- p %*% matrix.power(K, 1000)

pi
```

```
##           [,1]      [,2]      [,3]      [,4]      [,5]
## [1,] 0.148759 0.2352656 0.2634813 0.2097987 0.1426954
```

```r
# Validate global balance and stochasticity
Mod(pi %*% K - pi)
```

```
##      [,1]         [,2]         [,3]         [,4]         [,5]
## [1,]    0 2.775558e-17 5.551115e-17 2.775558e-17 2.775558e-17
```

```r
sum(pi)
```

```
## [1] 1
```

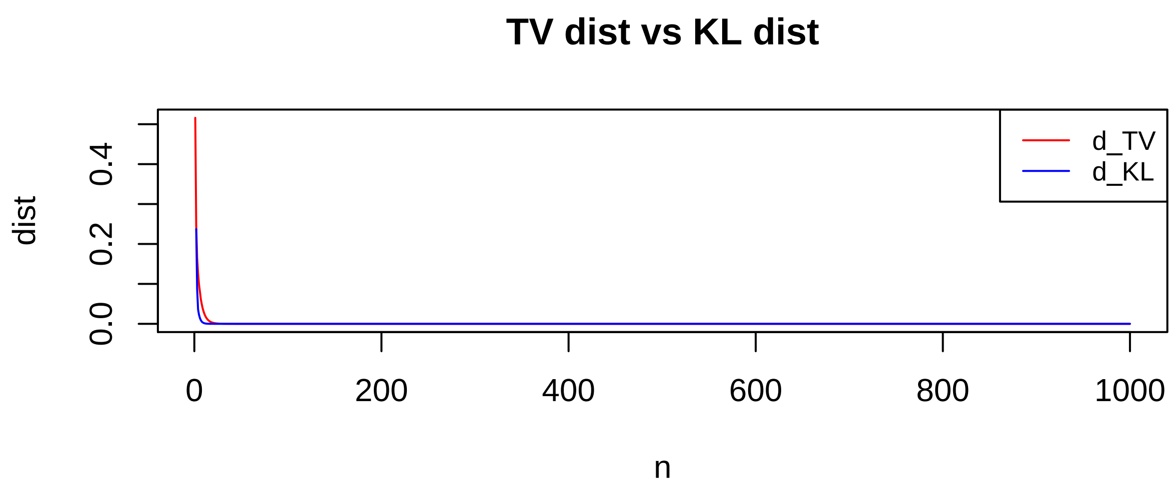
c)

```r
lambda_slem <- Mod(eigvals)[2]

lambda_slem
```

```
## [1] 0.7833305
```

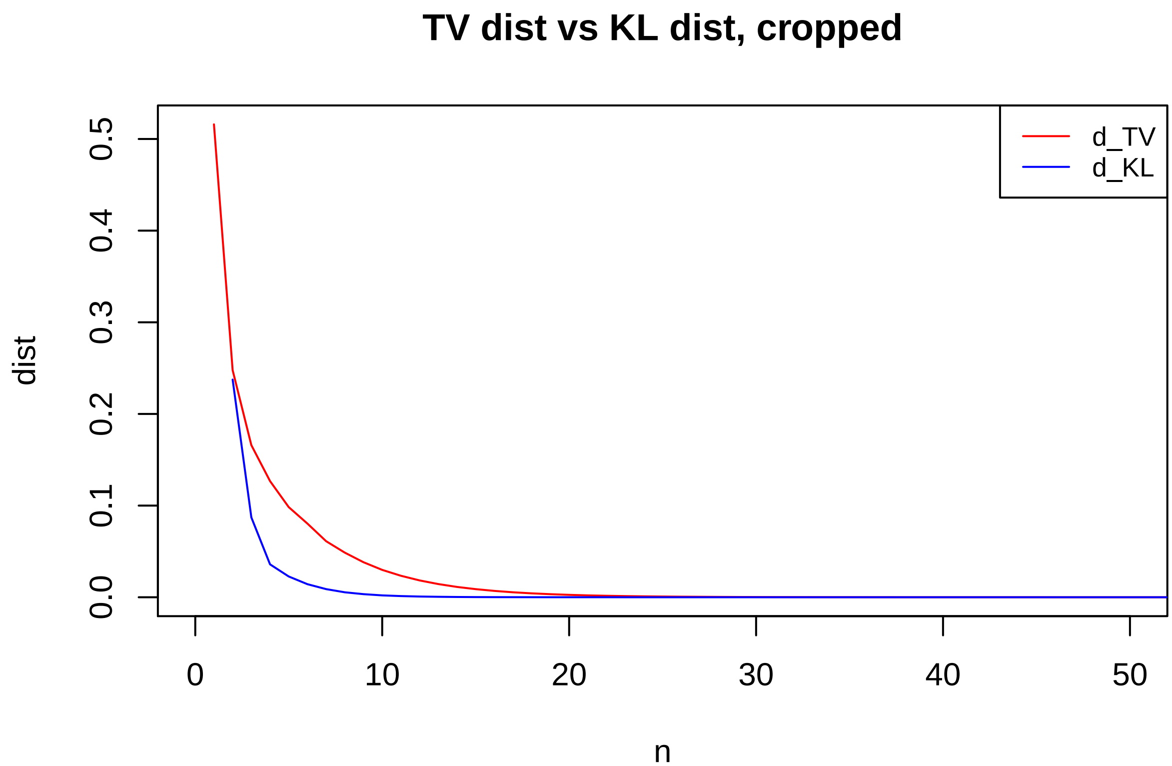
1.2.

```r
d_TV <- function(x) {
  1/2 * sum(abs(x))
}
d_KL <- function(pi, v) {
    sum(pi*log(pi / v))
}

v_0 <- c(1,0,0,0,0)

KK <- K

dists <- data.frame()

for (n in 1:1000) {
  v_n <- v_0 %*% KK
  dists <- rbind(dists, list(n = n, d_TV = d_TV(pi - v_n), d_KL = d_KL(pi, v_n)))
  KK <- KK %*% K
}

plot(dists$n, dists$d_TV, type="l", col = "red",
     main = "TV dist vs KL dist", xlab = "n", ylab = "dist")
lines(dists$n, dists$d_KL, col = "blue")

legend("topright", legend=c("d_TV", "d_KL"), col=c("red", "blue"), lty=1, cex=0.8)
```



Bounds are essentially the same after $n \approx 50$, hence we can "zoom" into this region to see more clearly.

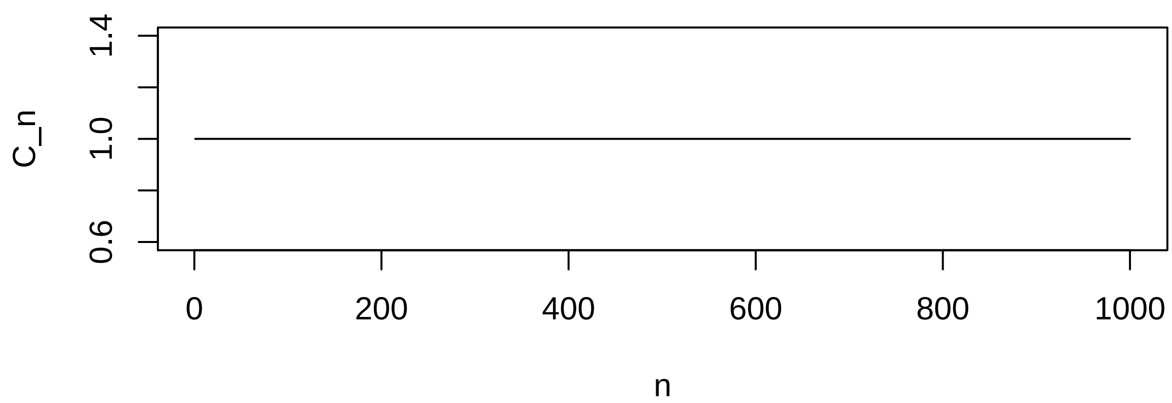
```r
plot(dists$n, dists$d_TV, type="l", col = "red", xlim=c(0,50),
     main = "TV dist vs KL dist, cropped", xlab = "n", ylab = "dist")
lines(dists$n, dists$d_KL, col = "blue")
legend("topright", legend=c("d_TV", "d_KL"), col=c("red", "blue"), lty=1, cex=0.8)
```



d_{KL} starts at infinity, as some elements are 0 in the first vectors ν_n for small n.

1.3.

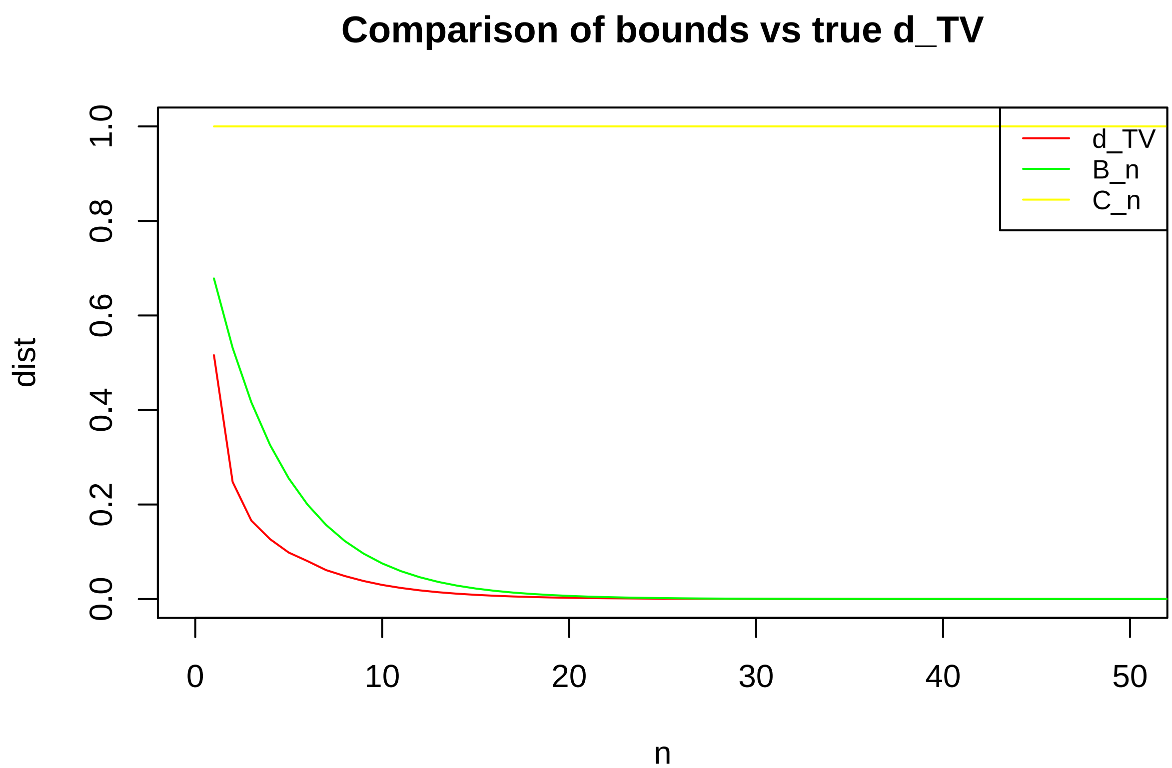
```r
combinations <- combn(5, 2)
row_dists <- c()

for (k in 1:ncol(combinations)) {
  x <- combinations[1, k]
  y <- combinations[2, k]
  row_dists <- c(row_dists, d_TV(K[x, ] - K[y, ]))
}

n <- 1:1000
C_n <- max(row_dists)^n

plot(n, C_n, type="l")
```



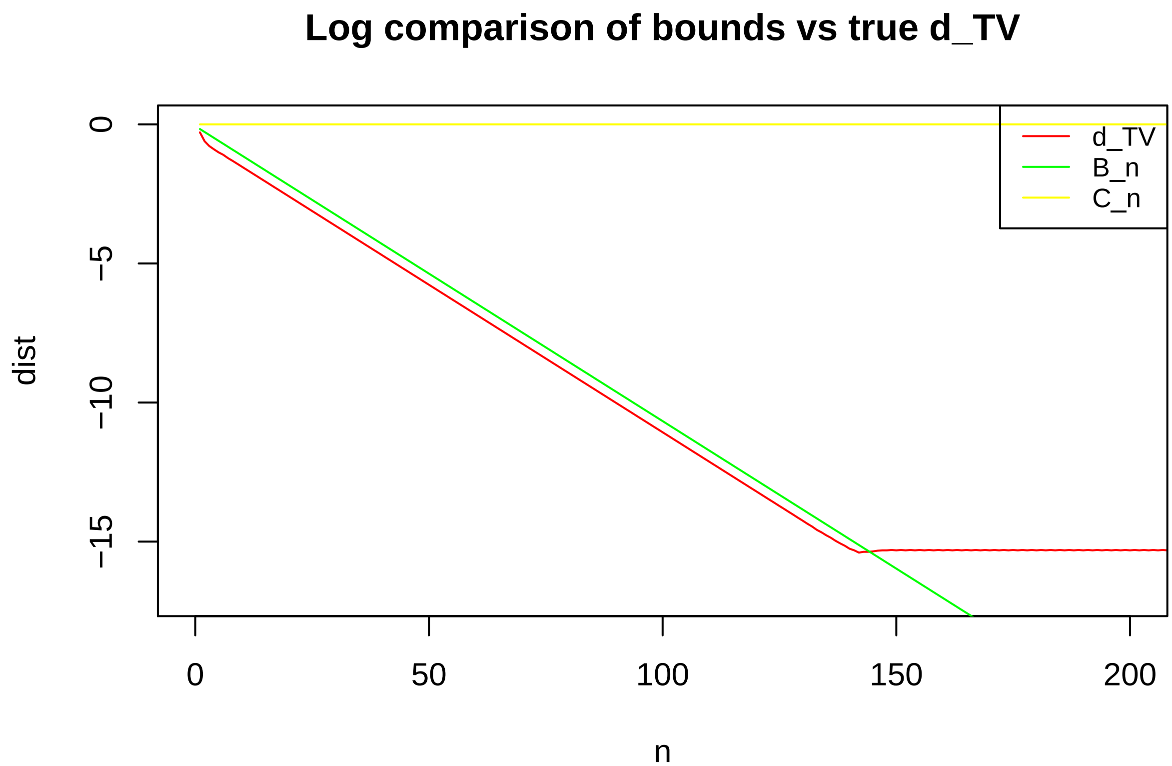
In this case, $C(K) = 1$, which means $C^n(K) = 1 \; \forall n$.

1.4.

```
pi_x0 <- pi[1]

B_n <- sqrt(1-pi_x0 / (4*pi_x0)) * lambda_slem^n

plot(dists$n, dists$d_TV, type="l", col = "red", ylim=c(0,1), xlim=c(0,50), main="Comparison of bounds v
lines(n, B_n, col = "green")
lines(n, C_n, col = "yellow")
legend("topright", legend=c("d_TV", "B_n", "C_n"), col=c("red", "green", "yellow"), lty=1, cex=0.8)
```



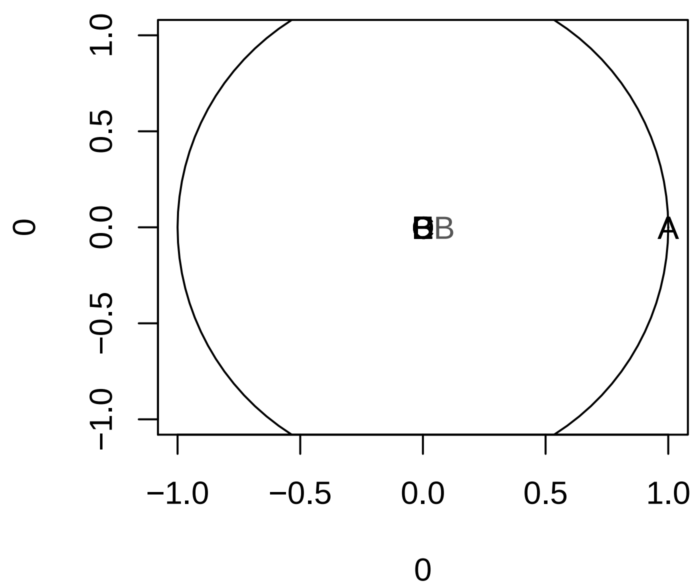

```
plot(dists$n, log10(dists$d_TV), type="l", col = "red", ylim=c(-17, 0), xlim=c(0,200), main="Log compari
lines(n, log10(B_n), col = "green")
lines(n, log10(C_n), col = "yellow")
legend("topright", legend=c("d_TV", "B_n", "C_n"), col=c("red", "green", "yellow"), lty=1, cex=0.8)
```



Log comparison of bounds vs true d_TV
d_TV
B_n
C_n
dist
0
−5
−10
−15
n
0
50
100
150
200

1.5.

```
ns <- c(10, 100, 1000)

eigs <- matrix(ncol = ncol(K), nrow=length(ns))

plot(0,0, ylim = c(-1, 1), xlim = c(-1, 1), type="n")

for (i in 1:length(ns)) {
  P <- matrix.power(K, ns[i])

  eigs[i, ] <- eigen(P)$values

  points(eigs[i,], pch=65:70, col = rgb(0,0,0, (i / 3)/2 + 0.5))

}

draw.circle(0, 0, 1)
```



Largest eig value stays on 1 + 0i, all other eigenvalues approach 0.

```r
P
```

```
##           [,1]      [,2]      [,3]      [,4]      [,5]
## [1,] 0.148759 0.2352656 0.2634813 0.2097987 0.1426954
## [2,] 0.148759 0.2352656 0.2634813 0.2097987 0.1426954
## [3,] 0.148759 0.2352656 0.2634813 0.2097987 0.1426954
## [4,] 0.148759 0.2352656 0.2634813 0.2097987 0.1426954
## [5,] 0.148759 0.2352656 0.2634813 0.2097987 0.1426954
```

```r
pi
```

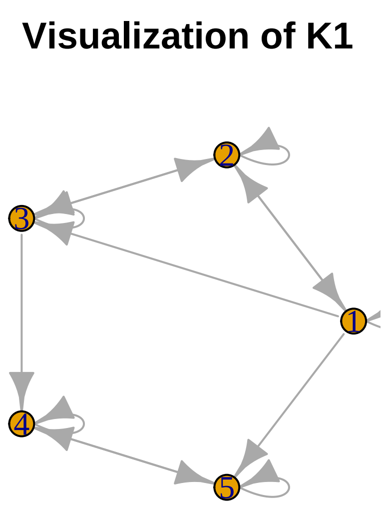
```
##           [,1]      [,2]      [,3]      [,4]      [,5]
## [1,] 0.148759 0.2352656 0.2634813 0.2097987 0.1426954
```

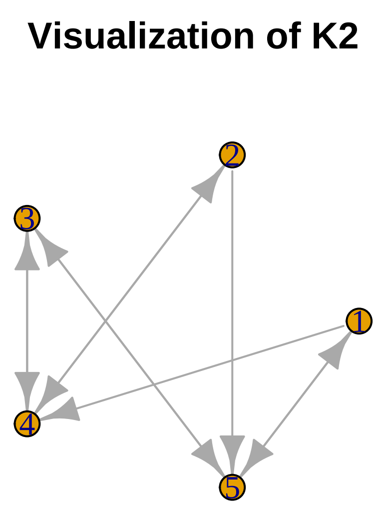
Yes it is “ideal” as it has π as rows.

2.

2.1.

```r
library(igraph)

K1 <- t(matrix(c(
  0.1, 0.4, 0.3, 0.0, 0.2,
  0.5, 0.3, 0.2, 0.0, 0.0,
  0.0, 0.4, 0.5, 0.1, 0.0,
  0.0, 0.0, 0.0, 0.5, 0.5,
  0.0, 0.0, 0.0, 0.7, 0.3
), nrow=5, ncol=5))


K2 <- t(matrix(c(
0.0, 0.0, 0.0, 0.4, 0.6,
0.0, 0.0, 0.0, 0.5, 0.5,
0.0, 0.0, 0.0, 0.9, 0.1,
0.0, 0.2, 0.8, 0.0, 0.0,
0.3, 0.0, 0.7, 0.0, 0.0
), nrow=5, ncol=5))

par(mfrow=c(1,2))

plot(graph_from_adjacency_matrix(K1 > 0), main="Visualization of K1", layout=layout_in_circle)
plot(graph_from_adjacency_matrix(K2 > 0), main="Visualization of K2", layout=layout_in_circle)
```





Above graph is a visualization, with connections for transition probabilities > 0, to illustrate the chains.

K_1 is not irreducible, as states 4 and 5 are accessible from all states, but states 1, 2 and 3 are not accessible from states 4, and 5. As such, there are two communication classes $\{1, 2, 3\}$ and $\{4, 5\}$, and the Markov Chain is reducible. On the other hand it is aperiodic.

K_2 is not aperiodic. This can be seen in the matrix, as it forms two "blocks". States 1, 2 and 3 always cycles to states 4 and 5, while states 4 and 5 always cycles to states 1, 2 and 3, implying a period of 2. On the other hand, all states are accessible, hence it is irreducible.

2.2.

Showing the left-eigenvalues as these include the invariant probabilities

```r
eigen(t(K1))
```

```
## eigen() decomposition
## $values
## [1]  1.0000000  0.9152351 -0.2000000 -0.1976553  0.1824202
##
## $vectors
##               [,1]       [,2]          [,3]         [,4]       [,5]
## [1,] -1.679212e-15 -0.2599052 -4.461753e-14  0.013002005 -0.6322914
## [2,] -3.108624e-15 -0.4237678  2.660802e-14 -0.007740231 -0.1042271
## [3,] -2.720046e-15 -0.3918867  1.152855e-14 -0.003372089  0.6629289
## [4,] -8.137335e-01  0.6399576  7.071068e-01 -0.707966017  0.3080528
## [5,] -5.812382e-01  0.4356022 -7.071068e-01  0.706076331 -0.2344632
```

```r
eigen(t(K2))
```

```
## eigen() decomposition
## $values
## [1] -1.000000e+00  1.000000e+00 -2.645751e-01  2.645751e-01 -3.270329e-18
##
## $vectors
##             [,1]        [,2]       [,3]       [,4]          [,5]
## [1,]  0.05041246 -0.05041246  0.5669467  0.5669467  6.172134e-01
## [2,]  0.14003460 -0.14003460 -0.3779645 -0.3779645 -7.715167e-01
## [3,]  0.67776746 -0.67776746 -0.1889822 -0.1889822  1.543033e-01
## [4,] -0.70017299 -0.70017299  0.5000000 -0.5000000  1.762007e-16
## [5,] -0.16804152 -0.16804152 -0.5000000  0.5000000 -1.180491e-16
```

2.3.

As matrices K_1 and K_2 are not primitive, the Perron-Frobenius Theorem does not hold. And the invariant probabilities can not be given from the eigenvectors directly, however they reveal something about them.

K_1 only has one eigenvalue $= 1$, this probability is

```
eigen(t(K1))$vectors[,1] / sum(eigen(t(K1))$vectors[,1])
```

```
## [1] 1.203761e-15 2.228450e-15 1.949894e-15 5.833333e-01 4.166667e-01
```

i.e a zero probability of being in states 1-3, and a non-zero probability of states 4 and 5. This is because 4 and 5 are accessible from 1-3, but not the other way. As such, as $n \to \infty$ there is a 0 probability of being in states 1-3, as all the mass will have moved to 4 and 5 and not been able to "escape".

We can see that the probability given by the largest eigenvalue eigenvectors is "stable", and hence the invariant probability of the chain.

```
(eigen(t(K1))$vectors[,1] / sum(eigen(t(K1))$vectors[,1])) %*% matrix.power(K1, 1000)
```

```
##              [,1]         [,2]         [,3]      [,4]      [,5]
## [1,] 4.447366e-54 7.251297e-54 6.705766e-54 0.5833333 0.4166667
```

K_2 is periodic and hence does not have a unique invariant probability, as $\lim_{n\to\infty} \nu_0 K_2$ does not converge for all ν_0. Rather, depending on its initial state, it either converges or cycles between two probabilities.

The stationary distribution, that fullfils $\pi K_2 = \pi$, is:

```
# Get from eigenvector
(pi <- (eigen(t(K2))$vectors[,2] / sum(eigen(t(K2))$vectors[,2])))
```

```
## [1] 0.02903226 0.08064516 0.39032258 0.40322581 0.09677419
```

```
# Multiply by K_2^1000 to show its stable
(pi <- pi %*% matrix.power(K2, 1000))
```

```
##            [,1]       [,2]      [,3]      [,4]       [,5]
## [1,] 0.02903226 0.08064516 0.3903226 0.4032258 0.09677419
```

```
# Show it satisfies global balance
pi %*% K2 - pi
```

```
##              [,1]         [,2]         [,3]          [,4] [,5]
## [1,] 3.469447e-18 1.387779e-17 1.110223e-16 -1.110223e-16    0
```

Example of initial state that "cycles" between probabilities, and does not solve $\pi K_2 = \pi$:

```
(pi <- c(1,0,0,0,0) %*% matrix.power(K2, 1000))
```

```
##            [,1]      [,2]      [,3] [,4] [,5]
## [1,] 0.05806452 0.1612903 0.7806452    0    0
```

```r
# Show it does not satisfy global balance
pi %*% K2 - pi
```

```
##              [,1]       [,2]       [,3]      [,4]      [,5]
## [1,] -0.05806452 -0.1612903 -0.7806452 0.8064516 0.1935484
```

```r
# Multiplying with K_2 again results in new distribution
pi %*% K2
```

```
##      [,1] [,2] [,3]      [,4]      [,5]
## [1,]    0    0    0 0.8064516 0.1935484
```

```r
# Multiplying again cycles back to the previous one
pi %*% K2 %*% K2
```

```
##             [,1]      [,2]      [,3] [,4] [,5]
## [1,] 0.05806452 0.1612903 0.7806452    0    0
```

3.

We have $P(\tau(0) = 1) = 1 - \alpha, P(\tau(0) = 2) = \alpha(1 - \alpha), P(\tau(0) = 3) = \alpha^2(1 - \alpha)$ and so on, as such $P(\tau(0) = k) = \alpha^{k-1}(1 - \alpha)\ \forall k > 0$. Finally, by the geometric sum formula:

$$\begin{aligned}
P(\tau_{\text{ret}}(0) < \infty) =& \sum_{k=1}^{\infty} P(\tau(0) = k) \\
=& \sum_{k=1}^{\infty} \alpha^{k-1}(1 - \alpha) \\
=&(1 - \alpha) \sum_{k=0}^{\infty} \alpha^k \\
\overset{0 \leq \alpha < 1}{=}&(1 - \alpha) \frac{1}{1 - \alpha} \\
=&1
\end{aligned}$$

For $\alpha = 1$, the walk deterministically goes up at every step and will never return to 0, hence a 0 probability. In summary, as long as $\alpha < 1$, the walk will return to state 0 with a 100% probability. In conclusion, 0 is a reccurent state for $\alpha < 1$.

For $\mathbb{E}(\tau_{\text{ret}}(0))$, we have for $0 \leq \alpha < 1$:

$$\begin{aligned}
\mathbb{E}(\tau_{\text{ret}}(0)) &= \sum_{k=1}^{\infty} k\alpha^{k-1}(1 - \alpha) \\
&= \frac{(1 - \alpha)}{\alpha} \sum_{k=1}^{\infty} k\alpha^k \\
&= \frac{(1 - \alpha)}{\alpha} \frac{\alpha}{(1 - \alpha)^2} \\
&= \frac{1}{1 - \alpha}
\end{aligned}$$

As $\sum_{k=1}^{\infty} kz^k = \frac{\alpha}{(1-z)^2}$ for $0 \leq z < 1$. For $\alpha = 1$, $\mathbb{E}(\tau_{\text{ret}}(0)) = 0$. In conclusion, 0 is a positively recurrent state for $\alpha < 1$

4.

Step 1

I will first show that

$$\mathrm{KL}_{X,Y}(\pi(X)P(X,Y)||\eta(X)P(X,Y)) = \mathrm{KL}_X(\pi(X)||\eta(X))$$

for any distribution η.

We have

$$\begin{aligned}
&\mathrm{KL}_{X,Y}(\pi(X)P(X,Y)||\eta(X)P(X,Y)) \\
&= \sum_{i,j} \pi(i)P(i,j) \log \frac{\pi(i)P(i,j)}{\eta(i)P(i,j)} \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} \pi(i)P(i,j) \log \frac{\pi(i)}{\eta(i)}
\end{aligned}$$

By P satisfying the detailed balance condition $\pi(i)K_{ij} = \pi(j)K_{ji}, \forall i, j \in \Omega$, and that implying global balance $\sum_{j=1}^{n} \pi(j)P(j,i) = \pi(i)$, we can simply the expression further as such:

$$\begin{aligned}
&\sum_{i=1}^{n}\sum_{j=1}^{n} \pi(i)P(i,j) \log \frac{\pi(i)}{\eta(i)} \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} \pi(j)P(j,i) \log \frac{\pi(i)}{\eta(i)} \\
&= \sum_{i=1}^{n} \log \frac{\pi(i)}{\eta(i)} w \\
&= \sum_{i=1}^{n} \log \frac{\pi(i)}{\eta(i)} \pi(i) \\
&= \mathrm{KL}_X(\pi(X)||\eta(X))
\end{aligned}$$

Step 2

Next, I will show that

$$\mathrm{KL}(\pi||\mu) - \mathrm{KL}(\pi||\nu) = \mathbb{E}_{Y\sim\pi}[\mathrm{KL}_X(P(Y,X)||Q(Y,X))]$$

where

$$Q(Y,X) = \frac{\mu(X)P(X,Y)}{\nu(Y)}$$

Expanding $\mathrm{KL}(\pi||\mu) - \mathrm{KL}(\pi||\nu)$, by using the detailed balance assumption, we get:

$$\begin{aligned}
&\mathrm{KL}(\pi||\mu) - \mathrm{KL}(\pi||\nu) \\
&= \sum_{i=1}^{n} \pi(i) \log \frac{\pi(i)}{\mu(i)} - \sum_{i=1}^{n} \pi(i) \log \frac{\pi(i)}{\nu(i)} \\
&= \sum_{i=1}^{n} \pi(i) [\log \frac{\pi(i)}{\mu(i)} - \log \frac{\pi(i)}{\nu(i)}] \\
&= \sum_{i=1}^{n} \pi(i) \log \frac{\nu(i)}{\mu(i)} \\
&= \sum_{i=1}^{n} \sum_{j=1}^{n} \pi(j) P(j,i) \log \frac{\nu(i)}{\mu(i)}
\end{aligned}$$

At the same time, expanding $\mathbb{E}_{Y\sim\pi}[\mathrm{KL}_X(P(Y,X)||Q(Y,X))]$, gives:

$$\begin{aligned}
&\mathbb{E}_{Y\sim\pi}[\mathrm{KL}_X(P(Y,X)||Q(Y,X))] \\
&= \sum_{i=1}^{n} \pi(i) \mathrm{KL}_X(P(i,X)||Q(i,X)) \\
&= \sum_{i=1}^{n} \pi(i) \sum_{j=1}^{n} P(i,j) \log \frac{P(i,j)}{Q(i,j)} \\
&= \sum_{i=1}^{n} \sum_{j=1}^{n} \pi(j) P(j,i) \log \frac{P(j,i)}{Q(j,i)} \\
&= \sum_{i=1}^{n} \sum_{j=1}^{n} \pi(j) P(j,i) \log \frac{P(j,i)}{\frac{\mu(i)P(i,j)}{\nu(j)}} \\
&= \sum_{i=1}^{n} \sum_{j=1}^{n} \pi(j) P(j,i) \log \frac{\nu(j)P(j,i)}{\mu(i)P(i,j)} \\
&= \sum_{i=1}^{n} \sum_{j=1}^{n} \pi(j) P(j,i) \log \frac{\sum_{k=1}^{n} \mu(k)P(k,j)P(j,i)}{\mu(i)P(i,j)}
\end{aligned}$$

Which gives a similar result as the previous expression, except the ν and μ have different indices, and there are P factors in the fraction. I have unfortunately not been able to finish this step, I believe the key to solve it is using $\nu = \mu P$.

Step 3

Assuming we have shown $\mathrm{KL}(\pi||\mu) - \mathrm{KL}(\pi||\nu) = \mathbb{E}_{Y\sim\pi}[\mathrm{KL}_X(P(Y,X)||Q(Y,X))]$, we need to show that $\mathbb{E}_{Y\sim\pi}[\mathrm{KL}_X(P(Y,X)||Q(Y,X))] \geq 0 \implies \mathrm{KL}(\pi||\nu) \leq \mathrm{KL}(\pi||\mu)$. This follows from KL being a measure of distance and is hence always ≥ 0 for stochastic vectors, and the probabilities in the sum forming the expected value always being ≥ 0, implying $\mathbb{E}_{Y\sim\pi}[\mathrm{KL}_X(P(Y,X)||Q(Y,X))] \geq 0$.